



# Contributions à l'élaboration de connaissances qualitatives en bio-informatique

Morgan Magnin

## ► To cite this version:

Morgan Magnin. Contributions à l'élaboration de connaissances qualitatives en bio-informatique. Bio-informatique [q-bio.QM]. Université de Nantes, 2016. tel-01321311

**HAL Id: tel-01321311**

**<https://theses.hal.science/tel-01321311>**

Submitted on 25 May 2016

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution| 4.0 International License

# Habilitation à Diriger des Recherches

Morgan MAGNIN

*Mémoire présenté en vue de l'obtention de  
l'habilitation à diriger des recherches de l'Université de Nantes  
sous le sceau de l'Université Bretagne Loire*

École doctorale : Sciences et technologies de l'information et mathématiques

Discipline : Informatique et applications

Unité de recherche : Institut de Recherche en Communications et Cybernétique de Nantes (IRCCyN)

Soutenue le 28 avril 2016

## Contribution à l'élaboration de connaissances qualitatives en bio-informatique

### JURY

Présidente :	<b>M<sup>me</sup> Mireille RÉGNIER</b> , Directrice de recherche Inria, Inria Paris-Saclay & École polytechnique (LIX)
Rapporteurs :	<b>M. Gilles BERNOT</b> , Professeur des universités, Université de Nice – Sophia Antipolis (I3S) <b>M. François FAGES</b> , Directeur de recherche Inria, Inria Paris-Saclay <b>M. Cédric LHOSSAINE</b> , Professeur des universités, Université de Lille 1 (LIFL)
Examineurs :	<b>M. Jérémie BOURDON</b> , Professeur des universités, Université de Nantes (LINA) <b>M. Hidde DE JONG</b> , Directeur de recherche Inria, Inria Grenoble - Rhône-Alpes <b>M. Denis THIEFFRY</b> , Professeur des universités, École normale supérieure (Institut de biologie)
Directeur de recherche :	<b>M. Olivier F. ROUX</b> , Professeur des universités, École centrale de Nantes (IRCCyN)



## REMERCIEMENTS

Cette HDR est le résultat d'un cheminement, tant personnel que scientifique, et il me tient à cœur, dans cette partie préliminaire, de remercier celles et ceux qui ont joué un rôle crucial dans mon parcours depuis ma thèse.

En premier lieu, mon épouse, Alice, qui m'a soutenu et encouragé tout au long de ces années. Un exemple parmi des millions d'autres : la collaboration avec le Japon, dont je rends compte dans ce manuscrit, n'aurait jamais pu se concrétiser si elle n'avait pas adhéré à ce projet au point de m'accompagner dans mes différents séjours. Je lui suis infiniment reconnaissant de sa bienveillance, tout comme je suis fier de voir combien elle parvient, au quotidien, à partager sa passion pour la culture japonaise à travers ses créations. J'invite vivement le lecteur de ce manuscrit à aller s'émerveiller sur le site web de Rosalys pour en découvrir plus : <http://www.rosalys.net>

Je souhaite aussi remercier mon ami d'enfance Sébastien, sa compagne Aude et leurs deux enfants Antoine et Jeanne pour leur amitié, aussi chaleureuse que le soleil du Sud !

Une HDR est tout autant un accomplissement qu'une étape dans la carrière d'un enseignant-chercheur. Je dois énormément aux collègues, en France ou au Japon, en recherche, enseignement ou TICE, avec lesquels j'ai eu la chance de travailler. L'ambiance chaleureuse, la bonne humeur et l'honnêteté sont autant de facteurs précieux qui donnent le sourire le matin en allant travailler et poussent à donner son maximum.

Merci à Olivier F. Roux, qui m'a accueilli avec beaucoup d'enthousiasme en 2008 en me proposant, dès mon arrivée, de co-encadrer un doctorant. On sait que les premières années de maître de conférences sont particulièrement chargées et qu'il n'est alors pas aisé de maintenir son lien avec la recherche. À cette occasion, et à bien d'autres, j'ai pu mesurer combien Olivier était un excellent chef d'équipe, mais surtout une personne précieuse.

Merci à chacun des étudiants de Master, doctorants et/ou post-doctorants que j'ai co-encadrés : Mylène Maurin, Bertrand Miannay, Anne-Laure Kabalu, Loïc Paulevé, Maxime Folschette, Courtney Chancellor, Simon Carolan, Emna Ben Abdallah, Xinwei Chai et Tony Ribeiro. Les échanges avec chacun d'entre eux ont été riches, non seulement sur le plan scientifique, mais également extra-professionnel. La recherche est un vrai travail d'équipe, et je saisis ma chance d'avoir rencontré des jeunes chercheurs aussi impliqués, talentueux et sympathiques, chacun dans leur style, dans leur personnalité.

Merci à Benjamin Vialle et Nelle Varoquaux pour l'aventure MarkUs, les suivis de projets

---

dans la bonne humeur, les déplacements ensemble à des événements autour des TICE et des logiciels libres.

Merci à Guillaume Moreau, qui m'a donné plusieurs belles opportunités, par exemple en 2008 en m'invitant à co-rédiger puis co-animer un projet TICE avec HP ou, en 2010, en permettant que je sois nommé chargé de mission sur ce sujet.

Merci à Patrice Cartraud, qui a succédé à Guillaume en tant que responsable direct sur les questions TICE. Nos échanges m'ont beaucoup appris, et j'ai apprécié son soutien sans faille, son efficacité et sa droiture.

Merci à Katsumi Inoue de m'avoir fait confiance quand, en 2011, je l'ai contacté pour évoquer des collaborations possibles, puis pour m'avoir donné l'opportunité de ces fructueux séjours de recherche à Tôkyô. Vivre le quotidien d'un autre laboratoire dans un pays tel que le Japon et explorer de nouvelles thématiques, voilà qui est une vraie chance. Merci également aux autres collègues japonais que j'ai rencontrés durant mes séjours, Mutsunori Banbara, Takehide Soh, Yoshitaka Yamamoto, Naoyuki Tamura, Chiaki Sakama, Tatsuya Akutsu et Takeyuki Tamura pour leur sens de l'accueil et les expériences enrichissantes vécues ensemble ! J'en profite également pour remercier mes différents professeurs de japonais, notamment Yumi Yoshida, car l'apprentissage de cette langue a facilité mon intégration et mes échanges scientifiques.

Merci à Jean-Pierre Elloy pour les nombreux échanges qui comptent, que ces discussions soient scientifiques ou non. Merci pour m'avoir toujours incité à ne pas perdre de vue le cap de l'HDR.

Et, justement, un remerciement spécial à ceux qui ont contribué à ce que je passe à l'acte pour la rédaction du présent manuscrit : Jean-Paul Comet, Michel Rueher, Patrick Le Callet, Saïd Moussaoui.

Merci à Michel Malabre, Franck Plestan, Jean-Yves Hascoët et Arnaud Poitou pour avoir accepté et soutenu mes projets scientifiques entre France et Japon et ce, même s'ils m'éloignaient physiquement de l'IRCCyN et de l'École centrale de Nantes (heureusement, avec les TIC, la distance numérique est moindre désormais).

Merci à Henri Angelino pour avoir permis d'amorcer les échanges entre le NII et mon équipe de recherche.

Merci à Jean-Pierre Berthet et Gwen Le Mauff pour nos discussions passionnées (et passionnantes) autour des TICE. Malgré la distance, nous avons réussi à établir un pilier important de collaboration pour nos établissements.

Merci à Jean-Marie Gilliot, Anne-Céline Grolleau et Christine Vaufrey pour l'incroyable projet que représenta ITyPA.

Merci à Paco Chinesta pour son enthousiasme rayonnant et pour les discussions inspirantes tant sur les TICE que sur la bio-informatique.

Merci à tous les membres, passés et présents, de l'équipe MeForBio, qui ont concouru à sa très bonne ambiance et au plaisir d'échanger idées et conseils.

---

Merci à tous les membres du département TICE et Pédagogie pour ce que nous construisons ensemble, au fil des semaines, au fil des mois.

Merci à mes collègues du département Info-Maths pour leur intérêt pour l'enseignement, un trait que nous partageons. Je leur suis reconnaissant pour leur ouverture quand, plusieurs fois, j'ai débarqué avec un projet sous le bras, qu'il s'agisse des TCS, de MarkUs ou de l'approche orientée compétences. Je leur dois également d'avoir pallié à mon absence au cours de mon année de délégation CNRS.

Merci aux communautés des personnels de l'École centrale de Nantes, de l'IRCCyN et du NII, précieux creusets d'émulation intellectuelle.

Merci à chacun des membres de mon jury d'HDR : Cédric Lhoussaine, Denis Thieffry, François Fages, Gilles Bernot, Hidde de Jong, Jérémie Bourdon et Mireille Régner. Lorsque j'ai commencé à m'intéresser à la bio-informatique en 2008, à lire des publications ou prendre part à des écoles de jeunes chercheurs, leurs travaux ont été des sources d'inspiration. J'étais honoré que chacun d'entre eux accepte de prendre part à mon HDR. Leurs retours, que ce soit à travers les rapports sur mon manuscrit ou lors de la soutenance, sonnent à la fois comme des encouragements et comme des guides vers de nouvelles directions à explorer. Je garderai de la soutenance un inoubliable souvenir d'échanges scientifiques.

Enfin merci aux créateurs d'œuvres humanistes et positives, qu'il s'agisse de mangakas (tels Naoko Takeuchi et Hiro Mashima), de réalisateurs d'*anime* (à l'instar de Hayao Miyazaki et Mamoru Hosoda) ou plus généralement de bandes-dessinées (par exemple Geerts) et de séries (telles les équipes derrière Doctor Who), pour concourir à répandre des valeurs auxquels je crois profondément !



## TABLE DES MATIÈRES

	Page
<b>Table des figures</b>	<b>ix</b>
<b>1 CV détaillé</b>	<b>1</b>
1.1 Notice individuelle . . . . .	1
1.2 Encadrements et rayonnement . . . . .	7
1.2.1 Encadrements . . . . .	7
1.2.2 Collaborations scientifiques personnelles. . . . .	10
1.2.3 Activité d'évaluation de la recherche . . . . .	13
1.2.4 Activités d'organisation . . . . .	14
1.2.5 Soutiens financiers et valorisation - Projets et contrats . . . . .	14
1.2.6 Valorisation logicielle . . . . .	16
1.2.7 Cours et séminaires invités . . . . .	16
1.2.8 Médias . . . . .	19
1.2.9 E-learning . . . . .	19
1.3 Publications . . . . .	22
1.3.1 Revues internationales avec comité de lecture . . . . .	22
1.3.2 Revues nationales avec comité de lecture . . . . .	23
1.3.3 Chapitres de livres . . . . .	23
1.3.4 Conférences internationales avec comité de lecture et actes . . . . .	23
1.3.5 Conférences nationales avec comité de lecture et actes . . . . .	25
1.3.6 Workshops internationaux avec comité de lecture et actes . . . . .	26
<b>2 Introduction</b>	<b>27</b>
2.1 Les défis de la modélisation en biologie : une vision quantitative ou qualitative ? .	27
2.2 Le contexte historique : modélisation qualitative booléenne ou multi-valuée . . . .	29
2.3 Les défis de l'analyse . . . . .	30
2.4 Nos objectifs . . . . .	32
2.5 Organisation du document . . . . .	32
<b>3 Modéliser et abstraire</b>	<b>35</b>



3.1	Pourquoi abstraire ? . . . . .	35
3.2	Distance entre système et modèle . . . . .	36
3.3	Les enjeux de la modélisation en biologie . . . . .	38
3.4	Une large classe de modèles . . . . .	38
3.4.1	Un aperçu des principaux modèles existants . . . . .	39
3.4.2	Différence entre les modèles . . . . .	42
3.4.3	Comment choisir un modèle ? . . . . .	45
3.4.4	Limites des modèles existants . . . . .	47
3.4.5	Notre contribution : les Frappes de Processus (ou pourquoi un autre modèle) . . . . .	48
3.4.6	Les Frappes de Processus avec priorités . . . . .	61
3.5	Synthèse . . . . .	63
<b>4</b>	<b>Analyser</b>	<b>65</b>
4.1	Quelles sont les propriétés d'intérêt ? . . . . .	66
4.1.1	Recherche de points fixes . . . . .	66
4.1.2	Interprétation abstraite (accessibilités successives) . . . . .	67
4.1.3	Simulation stochastique . . . . .	69
4.2	Aperçu des résultats obtenus . . . . .	70
4.2.1	L'outil Pint . . . . .	70
4.2.2	Exemples biologiques . . . . .	71
4.3	Positionnement des Frappes de Processus par rapport aux formalismes existants . . . . .	76
4.3.1	Équivalences avec les réseaux booléens et le modèle de Thomas . . . . .	76
4.3.2	Équivalences entre les Frappes de Processus . . . . .	76
4.3.3	Équivalences avec les réseaux de Petri . . . . .	77
4.3.4	Équivalences complémentaires . . . . .	78
4.4	Synthèse . . . . .	78
<b>5</b>	<b>Inférer, voire apprendre (automatiquement)</b>	<b>81</b>
5.1	Approches logiques : induction, déduction, abduction et machine learning . . . . .	82
5.2	Inférence de paramètres discrets des réseaux de régulation biologiques . . . . .	82
5.3	Reconstruction d'un réseau de régulation (réseau Booléen) et de ses règles logiques . . . . .	84
5.3.1	Les enjeux de l'apprentissage . . . . .	84
5.3.2	Modéliser les influences avec retard en réseaux booléens . . . . .	85
5.3.3	Apprendre des réseaux booléens avec des influences retardées . . . . .	87
5.3.4	Données en entrée : traces d'exécution . . . . .	88
5.3.5	Apprentissage de réseaux booléens . . . . .	88
5.3.6	Benchmarks . . . . .	88
5.4	Promesses de l'apprentissage automatique pour la biologie . . . . .	90
5.5	Synthèse . . . . .	90

5.5.1	Inférence des relations logiques entre gènes . . . . .	90
5.5.2	Apprendre des influences retardées au sein de systèmes biologiques . . . .	91
<b>6</b>	<b>Conclusion générale</b>	<b>93</b>
6.1	Contribution . . . . .	93
6.2	Perspectives . . . . .	94
6.2.1	Apprentissage de réseaux de régulation biologiques modélisant des com- portements complexes . . . . .	95
6.2.2	Révision de modèles . . . . .	95
6.2.3	Étude de dynamiques complexes . . . . .	97
6.2.4	Champs d'application : vers la médecine personnalisée . . . . .	98
	<b>Bibliographie</b>	<b>101</b>



## TABLE DES FIGURES

FIGURE	Page
3.1 Exemple de Frappes de Processus standards. . . . .	52
3.2 Représentation de la dynamique du modèle de Frappes de Processus standards de la figure 3.1. . . . .	55
3.3 Exemple de Frappes de Processus avec une sorte coopérative. . . . .	58
3.4 Amélioration du modèle de Frappes de Processus de la figure 3.1. . . . .	58
3.5 Exemple de Frappes de Processus avec 3 classes de priorité. . . . .	63
4.1 Exemples d'intervalles de tir et de paramètres stochastiques. . . . .	75
4.2 Enrichissement des Frappes de Processus standards à l'aide de paramètres stochastiques	75
4.3 Fichier source pour le modèle de boucle incohérente de rétroaction décrite dans [118].	76
4.4 Synthèse des équivalences (en termes de (bi)simulation faible) entre les différentes Frappes de Processus. . . . .	77
4.5 Synthèse des équivalences (en termes de (bi)simulation faible) entre les Frappes de Processus et quelques-uns des principaux formalismes de la littérature. . . . .	78
5.1 Exemple d'une boucle de rétroaction . . . . .	85
5.2 Benchmarks illustrant les performances de l'algorithme LFkT . . . . .	88



## CV DÉTAILLÉ

## 1.1 Notice individuelle

### État civil

Morgan MAGNIN  
Né le 4 Décembre 1980.  
Marié.

### Coordonnées professionnelles actuelles

#### RECHERCHE

IRCCyN - UMR CNRS 6597  
1 rue de la Noë  
BP 92101  
44321 NANTES cedex 03

téléphone : 02 40 37 69 78  
fax : 02 40 37 69 30  
morgan.magnin@irccyn.ec-nantes.fr

#### ENSEIGNEMENT

École Centrale de Nantes  
1 rue de la Noë  
BP 92101  
44321 NANTES cedex 03

téléphone : 02 40 37 16 28  
morgan.magnin@ec-nantes.fr

## Parcours professionnel

- Depuis 2015*      **Nommé maître de conférences associé (*visiting associate professor*) au National Institute of Informatics (NII) de Tôkyô.** Position octroyant un financement pour des séjours de 2 mois chaque année.
- 2014-2015*      **Délégation CNRS.** Dix mois au National Institute of Informatics (Tôkyô, Japon) en tant que chercheur invité (sur un financement JSPS obtenu indépendamment) et deux mois à l'Institut de Recherche en Communications et Cybernétique de Nantes (IRCCyN).
- Depuis 2008*      **Maître de Conférences à l'École Centrale de Nantes.** Département Informatique et Mathématiques. Membre de l'équipe de recherche **MeForBio** (*Méthodes Formelles pour la Bio-informatique*) de l'IRCCyN, dirigée par Olivier F. Roux.
- 2007-2008*      **ATER (50%) à l'École Centrale de Nantes.** Département Informatique et Mathématiques.
- 2004-2007*      **Doctorant boursier CNRS BDI.** . Membre de l'équipe de recherche **Systèmes Temps Réel** de l'IRCCyN, dirigée par Yvon Trinquet, puis Olivier H. Roux<sup>1</sup>. **Moniteur à l'École Centrale de Nantes.** Département Informatique et Mathématiques.

---

1. Afin d'éviter toute confusion, il me faut mentionner que, même s'ils sont tous deux Professeurs des Universités à l'École Centrale de Nantes et tous deux responsables d'une équipe de l'IRCCyN (respectivement Systèmes Temps Réel et MeForBio), Olivier H. Roux et Olivier F. Roux sont bel et bien deux personnes différentes. Le premier fut mon directeur de thèse de doctorat, le second est le directeur de recherche de ce présent manuscrit d'HDR.

## Activités pédagogiques

- 2012 et 2013**      **MOOC connectiviste.** Co-créateur du 1<sup>er</sup> MOOC francophone, ITyPA ("Internet Tout y est Pour Apprendre"). Thème : *méthodes et outils pour apprendre sur Internet*. Durée : 10 semaines. Nombre d'inscriptions : 1300 inscrits en 2012, 2600 inscrits en 2013 dont environ 40 ont été certifiés (l'inscription à la certification n'était pas automatique).
- 2012**              **Au PRES UEB.** Enseignement des *usages pédagogiques des médias et réseaux sociaux dans le supérieur* : à un public de responsables TICE et ingénieurs pédagogiques (12h CM).
- 2011**              **À l'ESEO.** Enseignement des *enjeux légaux de la société de l'information et de la communication* (4h CM).
- Depuis 2004**      **À l'École Centrale Nantes.** Environ 260h TD par an depuis 2008-09, 96h TD en 2007-08 et 64h TD de 2004-05 à 2006-07.
- Responsable (*i.e.*, définition des contenus et production des supports associés) des enseignements suivants :
    - Depuis 2008-09 : *Bases de données et implantation en SQL* (L3, 180 étudiants par an, module de 4h CM + 8h TD + 12h TP)
    - Depuis 2008-09 : *Droit dans la société de l'information et de la communication* (M2, 40 étudiants par an, module de 12h CM)
    - De 2010-11 à 2013-14 : *Projets de veille technologique* (M2, 40 étudiants par an, module de 20h de travail par élève)
    - De 2009-10 à 2013-14 : *Méthodes et technologies web* (M1, cours en anglais, 48 étudiants par an, module de 16h CM + 4h TP)
    - De 2008-09 à 2013-14 : *Introduction aux systèmes d'exploitation* (M2, 40 étudiants par an, module de 9h CM + 3h TP)
  - Intervention en TD et en TP en tant que membre de l'équipe pédagogique des cours d'*algorithmique et programmation impérative en C* (L3) et de *programmation objet en Java* (M1)
  - Encadrement de *projets de développement* (M1 et M2) et de *travaux de fin d'étude* (M2)



### Compétences linguistiques

*Français*                      **Langue maternelle**

*Anglais*                      **Courant.** Titulaire du First Certificate in English of Cambridge University (2002).

*Allemand*                      **Maîtrise de la langue écrite, assez bonne maîtrise de la langue orale..** Titulaire du Zentrale Mittelstufen Prüfung de l'Institut Goethe (2002).

*Japonais*                      **Assez bonne maîtrise de la langue écrite et orale..** Titulaire du Japanese Language Proficiency Test N3 (2015), équivalent B1/B2 CECRL.

### Activités de recherche depuis la thèse

- Depuis 2015*      **En tant que maître de conférences associé (*visiting associate professor*) au National Institute of Informatics (NII) de Tôkyô**, séjours de 2 mois par année académique : pour 2015-2016, séjour en février et mars 2016.
- 2015*              **Chercheur invité (*visiting researcher*) au NII (Tôkyô)**. Séjour de 3 mois (juin-août 2015) dans le cadre du programme *MOU grant*.
- 2014-2015*      **Chercheur invité (*JSPS invited fellow*) au NII (Tôkyô)**. Séjour de 10 mois (juin 2014-mars 2015) dans le cadre du programme *JSPS FY2014 Long-term invitation fellowship*.
- 2013*              **Chercheur invité (*visiting researcher*) au NII (Tôkyô)**. Séjour de 3 mois (mars-mai 2013) dans le cadre du programme *Lecture series*.
- Depuis 2008*      **Maître de conférences membre de l'équipe *Méthodes Formelles pour la Bio-informatique (MeForBio)* (équipe créée en 2011, émanant de l'équipe-projet *Modélisation et Vérification des Systèmes Embarqués (MOVES)*) de l'IRCCyN**. Thèmes de recherche : conception de méthodes de vérification formelle pour la bio-informatique, incluant :
- Model-checking des réseaux de régulation biologiques
  - Inférence des paramètres des réseaux de régulation biologiques
  - Approches méthodologiques basées sur les complémentarités entre modèles (discrets, hybrides, stochastiques, etc.)
- 2007-2008*      **ATER membre de l'équipe *Systèmes Temps Réel* de l'IRCCyN**.

### Formation initiale

- 2004-2007*      **Doctorat de l'École Centrale Nantes**, spécialité "Automatique et Informatique Appliquée". Préparé à l'IRCCyN au sein de l'équipe *Temps Réel* et soutenu le 13 décembre 2007. Sous la direction de : Olivier H. Roux et Pierre Molinaro. Jury : Béatrice Bérard (prés.), Guy Juanole (rapp.), Patrice Moreaux (rapp.), Hanifa Boucheneb, Jean-Pierre Elloy (invité).
- 2003-2004*      **DEA "Automatique et Informatique" Appliquée de l'École Centrale de Nantes**. Filière "Système Temps Réel".
- 2001-2004*      **Diplôme d'Ingénieur de l'École Centrale de Nantes**. Option Informatique.

### Responsabilités collectives

- Depuis 2015*      **Directeur du département TICe & Pédagogie de l'École Centrale de Nantes**. Responsable d'une équipe de 5 personnes (2.88 ETP). Prolongement des activités précédemment menées en tant que chargé de mission. Membre de fait du comité de direction élargi de l'École Centrale de Nantes. Budget annuel moyen de fonctionnement : 45.000€
- Depuis 2014*      **Co-responsable du groupe de travail EAT-TICE pour le Groupe des Écoles Centrales (GEC)**. Animation du groupe via des réunions bimestrielles, établissement et suivi d'une feuille de route à 5 ans.
- 2010-2015*      **Chargé de mission EAT-TICE à l'École Centrale de Nantes**. Direction de la Formation. Membre de fait du comité de direction élargi de l'École Centrale de Nantes. Les activités menées dans ce cadre ont conduit l'établissement à recruter différents profils complémentaires et à en faire un département dédié en 2015. Budget annuel moyen de fonctionnement : 50.000€
- Depuis 2010*      **Implication dans la vie de l'établissement :**
- Depuis 2014 : Membre élu du Conseil d'Administration de l'École Centrale de Nantes
  - 2010-2014 : Membre élu du Conseil des Études de l'École Centrale de Nantes

## 1.2 Encadrements et rayonnement

Mes travaux résultent de collaborations régulières avec non seulement des étudiants de Master de Recherche et de Doctorats - qui ont co-signé les articles référencés dans ce manuscrit -, mais également de collègues à l'échelle nationale ou internationale.

### 1.2.1 Encadrements

#### 1.2.1.1 Encadrements de thèses de doctorat :

**Loïc Paulevé** (thèse commencée le 1<sup>er</sup> octobre 2008 et soutenue le 6 octobre 2011) sur la *modélisation, simulation et vérification des grands réseaux de régulation biologique*

- Encadrement : Morgan Magnin (50 %), Olivier F. Roux (50 %).
- Financement : contrat doctoral spécifique normalien
- Jury : Claude Jard (président), François Fages (rapporteur) Serge Haddad (rapporteur), Jérôme Feret (examineur), Jean-Phillipe Vert (examineur), Oliver F. Roux (directeur de thèse), Morgan Magnin (co-encadrant).
- Publications associées :
  - Sur le contenu de la thèse : 3 revues internationales (TCSB [139], IEEE TSE [140], MSCS [142]), 1 workshop international (SASB [138]).
  - Sur les collaborations postérieures à la thèse : 1 chapitre de livre ([135]), 2 revues internationales (TCS [67, 69]), 1 revue nationale (TSI [137]), 1 conférence internationale (CMSB [66]), 2 workshops internationaux ([65, 68]).
- Devenir du docteur : chercheur CNRS (CR2) dans l'équipe BioInfo du Laboratoire de Recherche en Informatique (LRI) à l'Université Paris-Sud depuis septembre 2013.

**Maxime Folschette** (thèse commencée le 1<sup>er</sup> octobre 2011 et soutenue le 8 octobre 2014) sur la *modélisation algébrique de l'évolution et de la dynamique multi-échelles des réseaux de régulation biologique*

- Encadrement : Morgan Magnin (50 %), Olivier F. Roux (50 %).
- Financement : contrat doctoral MESR.
- Cadre : projet "ANR-10-BLAN-0218 - BioTempo" : Langages, concepts de temps et modèles hybrides pour l'analyse de modèles incomplets en biologie moléculaire (2010-2014)
- Publications associées : 1 chapitre de livre ([135]), 2 revue internationale (TCS [67, 69]), 1 revue nationale (TSI [137]), 1 conférence internationale (CMSB [66]), 2 workshops internationaux ([65, 68]).
- Devenir du docteur : ATER à l'Université de Nice Sophia-Antipolis depuis septembre 2015, après un post-doc dans l'équipe "Formal Methods" à l'Université de Kassel (Allemagne).

**Simon (J.) Carolan** (thèse commencée le 1<sup>er</sup> octobre 2012, soutenance prévue le 20 juin 2016) sur l'*ingénierie ontologique pour la création et la gestion de ressources pédagogiques adaptatives*.

- Encadrement : Morgan Magnin (30 %), Guillaume Moreau (30 %) et Francisco Chinesta (40 %).
- Financement : allocation de recherche spécifique de l'École Centrale de Nantes sur projet transversal inter-laboratoires.
- Publications associées : 5 conférences internationales (EMOOCs [28], ICALT [25, 26], ESDA [58], DI [30]), 1 conférence nationale (TICE [29]), 1 workshop national ([27]).

**Courtney Chancellor** (thèse commencée le 1<sup>er</sup> octobre 2012, arrêtée à la demande de l'étudiante le 30 juin 2014 par souhait familial de rentrer de manière anticipée aux États-Unis) sur les *ponts entre les approches discrètes, continues et probabilistes pour l'analyse des réseaux de régulation biologiques*

- Encadrement : Morgan Magnin (30 %), Olivier F. Roux (30 %) et Francisco Chinesta (40 %)
- Financement : contrat doctoral lié à la chaire I.U.F. de Francisco Chinesta.
- Publications associées : 1 chapitre de livre ([135]), 1 conférence internationale (CMSB [31])
- Commentaire : même si la thèse a été interrompue, elle a permis d'amorcer la collaboration avec Francisco Chinesta et ses échanges se prolongent aujourd'hui activement [37].
- Devenir de la doctorante : ingénieur R&D à Epic, Wisconsin (États-Unis) depuis octobre 2014.

**Emna Ben Abdallah** (thèse commencée en septembre 2014) sur *l'analyse des caractéristiques comportementales et de l'impact de perturbations dans les réseaux de régulation biologiques*.

- Encadrement : Morgan Magnin (50 %), Olivier F. Roux (50 %).
- Financement : bourse Erasmus Mundus EU-METALIC.
- Publications associées : 1 conférence internationale (BIBM [17])

**Xinwei Chai** (thèse commencée en septembre 2015) sur *la révision de modèles biologiques discrets*.

- Encadrement : Morgan Magnin (50 %), Olivier F. Roux (50 %).
- Financement : bourse China Scholarship Council.

#### 1.2.1.2 Encadrements de Master :

- **Anne-Laure Kabalu**, *Ingénierie pédagogique pour l'enseignement à distance via une salle de télé-présence*  
Stage de M2. Soutenance prévue en septembre 2016.

Encadrement : Simon (J.) Carolan (50 %), Morgan Magnin (50 %).

- **Xinwei Chai**, *Complétion des Réseaux de Régulation – une contribution pour les Frappes de Processus*

M2 soutenu le 28 août 2015.

Encadrement : Morgan Magnin (50 %), Olivier F. Roux (50 %).

Devenir de l'étudiant : doctorant au sein de l'équipe MeForBio de l'IRCCyN.

- **Bertrand Miannay**, *Construction automatique de graphes et analyse par contraintes dans le cadre de l'étude du myélome multiple*

M2 soutenu le 11 septembre 2014.

Encadrement : Carito Guziolowski (40 %), Morgan Magnin (30 %) et Stéphane Minvielle (30 %).

Devenir de l'étudiant : doctorant au sein de l'équipe MeForBio de l'IRCCyN.

- **Anne-Laure Kabalu**, *Conception et implémentation d'un MOOC en école d'ingénieur*  
Stage de M1.

Encadrement : Simon (J.) Carolan (50 %), Morgan Magnin (50 %).

Publication associée : 1 conférence internationale (DI [30])

Devenir de l'étudiante : étudiante en Master 2 à l'Université de Poitiers.

- **Maxime Folschette**, *Application de la logique de Hoare pour la détermination des paramètres de délai dans un système de régulation génique*

M2 soutenu le 1<sup>er</sup> septembre 2011.

Encadrement : Morgan Magnin (50 %), Olivier F. Roux (50 %).

Devenir de l'étudiant : ATER à l'Université de Nice Sophia-Antipolis depuis septembre 2015.

- **Mylène Maurin**, *Contribution à l'analyse stochastique des systèmes évolutionnaires : Application à la régulation génique et utilisation du pi-calcul*

M2 soutenu le 25 juin 2009

Encadrement : Morgan Magnin (50 %), Olivier F. Roux (50 %).

Publication associée : 1 conférence internationale (BiCoB [121])

Devenir de l'étudiante : programmeuse biostatistique au sein de Novartis Pharma à Paris.

- **Gilles Benattar**, *Logiques temporelles dédiées à la vérification de réseaux de Petri à chronomètres en temps discret*

M2 soutenu le 18 septembre 2007.

Encadrement : Morgan Magnin (50 %), Olivier H. Roux (50 %).

Devenir de l'étudiant : ingénieur R&D au sein de ClearSy à Paris.

- **Charlotte Seidner**, *Mise en œuvre d'une traduction préservant la bisimulation d'un modèle d'architecture fonctionnelle de haut niveau vers les réseaux de Petri temporels*

M2 soutenu le 20 septembre 2006

Encadrement : Morgan Magnin (50 %), Olivier H. Roux (50 %).

Cadre : en partenariat avec l'entreprise Sodius.

Devenir de l'étudiante : maître de conférences à l'IUT de Nantes.

### 1.2.1.3 Autres encadrements :

Au cours de mes séjours au National Institute of Informatics, j'ai travaillé avec **Tony Ribeiro** (doctorant supervisé par Katsumi Inoue) sur plusieurs chapitres de sa thèse. Cette collaboration a été concrétisée par une publication en revue (Frontiers in Bioengineering and Biotechnology [151]) et un article court en conférence internationale [150]. J'ai ainsi participé dans la pratique à l'encadrement de thèse de Tony Ribeiro, qui a soutenu son doctorat le 14 juillet 2015 devant un jury dont j'étais membre. Depuis le 1er octobre 2015, Tony Ribeiro est post-doc au sein de l'équipe MeForBio à Nantes et j'encadre son travail dans ce cadre.

De plus, j'ai co-encadré avec Katsumi Inoue un stage doctoral de 5 mois (octobre 2014-février 2015) de **Alexander Andreychenko** (doctorant à l'Université de Saarland en Allemagne). Ce co-encadrement a donné lieu à **une publication en conférence internationale** avec comité de lecture et publication des actes [8].

Enfin, j'ai encadré (à 100 %) une ingénieure pédagogique pendant un an (**Anne-Céline Grolleau**, du 1<sup>er</sup> septembre 2012 au 12 octobre 2013), de nombreux élèves en contrats étudiants sur des missions liées au e-learning, ainsi que de nombreux travaux de fin d'étude de l'École Centrale de Nantes. Depuis le 1<sup>er</sup> septembre 2015 et la mise en place d'un département dédié au e-learning au sein de l'école Centrale de Nantes, j'encadre (à 100 %) le mi-temps de deux ingénieures pédagogiques (**Vanessa Le Garrec** et **Carole Moreau**) et le mi-temps d'un technicien audio-visuel (**Patrick Roustang**).

## 1.2.2 Collaborations scientifiques personnelles.

### 1.2.2.1 Collaborations nationales :

**Avec Francisco (Paco) Chinesta, titulaire d'une chaire EADS au GeM, laboratoire de Génie Mécanique (Nantes).** Cette collaboration, surprenante en apparence (Paco est un spécialiste de la mécanique), résulte d'une curiosité partagée pour l'apport du numérique dans la pédagogie d'une part (c'est ainsi que notre échange a débuté), de la biologie d'autre part (notre discussion ayant ainsi dérivé sur nos domaines de recherche respectifs). Il en résulte :

- Dans le premier domaine : une thèse en co-encadrement (thèse de Simon (J.) Carolan sur la conception d'un dispositif d'apprentissage augmenté pour les sciences de l'ingénieur).
- Dans le second domaine, des travaux en commun sur les rapprochements entre les méthodes continues et discrètes pour l'étude des réseaux de régulation génétique. Cette collaboration a conduit à une thèse en co-encadrement (thèse de Courtney Chancellor sur les liens entre les approches discrètes, stochastiques et continues pour l'étude des réseaux de régulation biologique) et aux publications suivantes : 1 revue internationale [37] et 1 conférence internationale [31].

**Avec Gilles Bernot, Jean-Paul Comet, Franck Delaunay et Adrien Richard (Université de Nice).** Nous avons travaillé ensemble dans le cadre du PEPH CirClock, visant à l'élaboration d'un modèle hybride pour l'analyse des mécanismes génétiques à l'origine du rythme circadien des mammifères. Cette collaboration s'est traduite par plusieurs réunions, et est appelée à se poursuivre et se concrétiser dans le contexte du projet ANR HyClock.

**Avec Jean-Marie Gilliot (Telecom Bretagne), Anne-Céline Grolleau et Christine Vau-frey.** Ensemble, nous avons lancé en 2012 le premier MOOC francophone, ITyPA, sur les méthodes et outils pour apprendre via Internet. Nous avons assuré l'animation de ce MOOC connectiviste pendant deux ans (le portage du projet a ensuite été repris par mon doctorant Simon Carolan) et participé à de nombreuses actions de valorisation des MOOCs en France et en Europe. Cette initiative a par ailleurs donné lieu aux publications suivantes : 1 conférence internationale [28], 2 conférences nationales [29, 75] et 1 workshop national [27].

### 1.2.2.2 Collaborations internationales :

**Avec Katsumi Inoue (National Institute of Informatics, Tôkyô).** J'ai initié cette collaboration en 2011 et l'ai, depuis, inscrite dans la durée. Nous travaillons ensemble sur les mérites respectifs et la complémentarité des méthodes de vérification formelle et des approches logiques pour étudier les systèmes dynamiques complexes, notamment ceux qui proviennent de la biologie des systèmes. D'abord, Maxime Folschette, doctorant sous ma supervision avec Olivier F. Roux, a fait un stage de 3 mois grâce à l'accord NII-AtlanSTIC du 26 février 2012 au 25 mai 2012. Après avoir défendu sa thèse, il a bénéficié d'un financement MOU grant pour un séjour de recherche de 1 mois du 21 octobre 2014 au 20 novembre 2014, lui permettant d'identifier de nouvelles collaborations.

J'ai été invité par le NII pour une série de conférences du 18 Mars 2013 au 10 mai 2013. Dans la continuité de ce séjour de recherche, j'ai soumis ma candidature au programme de recherches au Japon JSPS (long terme). Celle-ci a été acceptée et j'ai ainsi rejoint le laboratoire de Katsumi Inoue de juin 2014 à mars 2015 pour travailler sur les systèmes dynamiques. Grâce à cette subvention JSPS, un large éventail de nouvelles recherches conjointes ont émergé.



Pour continuer, Emna Ben Abdallah, doctorante en 1<sup>ère</sup> année sous ma supervision avec Olivier F. Roux, vient d'effectuer un stage de 3 mois du 1<sup>er</sup> février 2015 au 27 avril 2015. Elle s'est concentrée sur des algorithmes pour la réalisation de réseaux biologiques grâce à la programmation logique. Enfin, en retour, Tony Ribeiro (doctorant au NII/Sokendai, sous la supervision de Katsumi Inoue) a donné une conférence au séminaire de recherche de l'équipe MeForBio le 21 octobre 2013 sur *l'apprentissage de réseaux booléens de transitions d'interprétation*. Il a, à nouveau, rendu visite à l'équipe le 10 mars 2014 pour discuter d'autres collaborations.

Nous avons concrétisé ces échanges par la suite à travers des publications conjointes référencées ci-dessous et préparées au cours de mes séjours au National Institute of Informatics. En définitive, j'ai participé dans la pratique à l'encadrement de thèse de Tony Ribeiro, qui a soutenu son doctorat le 14 juillet 2015 devant un jury dont j'étais membre. Enfin, suite à ces travaux conjoints, nous lui avons proposé un post-doc au sein de l'équipe MeForBio à Nantes.

Jusqu'à maintenant, cette collaboration avec Katsumi Inoue et son équipe a conduit à :

- 3 publications dans des revues internationales [67, 151, 160]
- 2 publications dans des conférences internationales [8, 66]
- 1 publication dans une conférence nationale [159]
- 1 publication dans un workshop international [65]

Une liste exhaustive des réunions de recherches associées est disponible sur <http://morganmagnin.net/bioinformatics/>

**Avec différentes équipes de recherche japonaises.** Lors de mes séjours au Japon, j'ai pris part à plusieurs rencontres scientifiques avec des différentes équipes de chercheurs. Ces discussions en cours sont résumées comme suit (cette liste est toutefois non exhaustive) :

- Travaux avec Tatsuya Akutsu et Takeyuki Tamura (Université de Kyoto) à propos de la révision de réseaux biologiques : réunions annuelles KUBIC en 2013 et 2014, ainsi que deux journées dédiées au travail sur la complétion de réseau le 6 février (à Kyoto) et le 25 mars 2015 (à Tôkyô).
- Travaux avec Naoyuki Tamura, Mutsunori Banbara et Takehide Soh (Université de Kobe) autour d'algorithmes de model-checking implantés en Answer Set Programming : j'ai donné un séminaire à Kobe en août 2014, puis ai prolongé ces échanges lors d'un workshop à Ishigaki en mars 2015 et d'un nouveau séminaire à Kobe en août 2015.
- Discussions avec Chiaki Sakama (Université de Wakayama) à propos de l'abduction dans les réseaux biologiques [151]
- Travaux avec Yoshitaka Yamamoto (Université de Yamanashi) à propos des observations incohérentes avec des modèles exprimés en Frappes de Processus : participation conjointe aux séminaires LRI-NII-Yamanashi en 2013 et 2014, séminaire à Yamanashi en juin 2014, et multiples réunions conjointes.

- Travaux avec les membres du projet *Résilience des Systèmes* dirigé par Hiroshi Maruyama (<http://systemsresilience.org/index-e.html>) : réunions mensuelles, participation à un Shonan Meeting d'une semaine en février 2015, une publication en revue acceptée [160].

**Avec Karen Reid de l'Université de Toronto.** J'ai participé à la mise en place d'un partenariat informel entre l'École Centrale de Nantes et l'Université de Toronto autour du logiciel MarkUs, pour l'évaluation des travaux d'étudiants en travaux pratiques d'informatique. Cette collaboration a donné lieu aux publications suivantes : 1 publication dans une conférence internationale [117], deux publications dans des conférences nationales [114, 116]. De plus, MarkUs a reçu en 2011 la "mention spéciale" des 3<sup>èmes</sup> Trophées des Technologies Éducatives décernées au salon professionnel de l'éducation Educatec-Educative.

### 1.2.3 Activité d'évaluation de la recherche

#### 1.2.3.1 Comités de programme

J'ai été membre du comité de programme ("PC member") des conférences suivantes :

- IJCAI'2015 (International Joint Conference on Artificial Intelligence), Knowledge Representation and Reasoning (KR) Track, Buenos Aires, Juillet 2015.
- TICE'2014 ( Technologies de l'Information et de la Communication pour l'Enseignement), Comité "Retour d'Expériences", Béziers, Novembre 2014.
- Premières Rencontres des Jeunes Chercheurs en Informatique Temps Réel (RJCITR'05), Nancy, Septembre 2005.

#### 1.2.3.2 Relectures d'articles

J'ai effectué des évaluations d'articles ("review") pour :

- des revues internationales : MethodsX, IEEE TSE, Acta Cybernetica, TCSB ;
- des revues nationales : JESA, TSI ;
- des conférences internationales : PN'2016, AAAI'2015, PSB'2015, FMMD'2014, FRO-COS'2013, CP'2012, CMSB'2011, ICATPN'2010, FORMATS'2008 ;
- des conférences nationales : TICE'2014, JOBIM'2012.

#### 1.2.3.3 Participations à des commissions de recrutement

J'ai été membre des comités de sélection suivants :

- en tant que membre externe : poste bio-informatique "27 MCF 4237" ouvert en 2014 à l'Université de Nice.
- en tant que membre interne : poste bio-informatique "27 MCF 558" ouvert en 2012 à l'École Centrale de Nantes (chaire d'excellence CNRS) ;

### 1.2.3.4 Expertise de projets

J'ai effectué des évaluations de 5 projets soumis dans le cadre de l'"Appel à Projets 2011 dans le domaine de la physique, des mathématiques ou des sciences de l'ingénieur appliqués au Cancer" organisé par l'INSERM dans le contexte général du "Plan Cancer 2009-2013".

### 1.2.4 Activités d'organisation

J'ai été membre du comité d'organisation des événements suivants :

- 13<sup>th</sup> International Conference on Computational Methods in Systems Biology (CMSB 2015). 91 participants. Présidé par Jérémie Bourdon (LINA, Nantes) et Olivier F. Roux (IRCCyN, Nantes).
- 10<sup>e</sup> Journées Ouvertes en Biologie, Informatique et Mathématiques (JOBIM 2009). 350 participants. Manifestation co-présidée par Rémi Houlgatte (INSERM, Nantes) et Jean-Michel Richer (LERIA, Angers).
- 7<sup>e</sup> colloque francophone sur la Modélisation des Systèmes Réactifs (MSR 2009). 70 participants. Présidé par Olivier H. Roux (IRCCyN, Nantes).

### 1.2.5 Soutiens financiers et valorisation - Projets et contrats

Les déplacements liés à ces collaborations et à mes travaux ainsi que certains salaires des doctorants ont été possibles grâce à différents projets (un projet régional GRIOTE, deux projets ANR - BioTempo et HyClock - et trois projets japonais, deux projets UNIT, un projet avec l'entreprise HP) auxquels j'ai pris part. Cette section résume ces projets.

#### 1.2.5.1 En tant que porteur

- J'ai été porteur, de 2008 à 2010, du projet HP "Technology for Teaching" baptisé *Most complex systems are designed with the simplest tools*. Le but de ce projet était la conception de dispositifs pédagogiques innovants impliquant des tablettes tactiles. La dotation de ce projet, que j'avais co-rédigé avec Guillaume Moreau, était de 100.000 US\$. Ce travail a été synthétisé dans une publication en conférence nationale [112].
- Mon premier séjour au National Institute of Informatics (NII) de Tôkyô (3 mois, de février à mai 2013) a été pris en charge à 100% par le NII dans le cadre d'un poste de *chercheur conférencier*. Le budget alloué était de 550.000 JPY.
- Mon second séjour au National Institute of Informatics (NII) de Tôkyô (10 mois, du 1<sup>er</sup> juin 2014 au 31 mars 2015) a été pris en charge à 100% par un projet japonais (JSPS FY2014 Long-term invitation fellowship) que j'avais déposé avec mon chercheur hôte, Katsumi Inoue. Ce projet était doté d'un budget de 3.940.000 JPY.

- Mon troisième séjour au National Institute of Informatics (NII) de Tôkyô (3 mois, de juin à août 2015) a été pris en charge à 100% sous la forme d'un poste de *chercheur invité* (MOU grant). La dotation était de 1.000.000 JPY.
- Depuis avril 2015, je suis chercheur associé au au National Institute of Informatics (NII) de Tôkyô, ce qui me confère une dotation annuelle de fonctionnement de 400.000 JPY.

### 1.2.5.2 En tant que participant

Depuis ma thèse, j'ai pris part aux projets suivants.

#### Projets en cours

- Projet ANR HyClock (budget : 500.000€)  
Objet : Modélisation Hybride Formelle du Temps pour la Biologie des Horloges Circadiennes et la Chronopharmacologie  
Porteur : Franck Delaunay (IBV, Nice)  
Démarré en janvier 2015  
Durée : 4 ans
- Projet GRIOTE financé par la région "Pays de la Loire" (budget : 1.835.000€)  
Objet : *Groupeement de Recherche en Intégration de données Omics à Très grande Échelle* dédié au renforcement des liens entre les équipes de bio-informatique en Pays de la Loire  
Porteur : Jérémie Bourdon (LINA, Université de Nantes)  
Démarré en octobre 2013  
Durée : 4 ans

#### Projets terminés

- Projet ANR BioTempo (budget : 374.000€)  
Objet : Langages, concepts de temps et modèles hybrides pour l'analyse de modèles incomplets en biologie moléculaire  
Porteur : Anne Siegel (IRISA, Rennes)  
Démarré en mars 2011  
Durée : 3 ans et 6 mois
- Projet CNRS PEP2 CirClock (budget : 20.000€)  
Objet : Modélisation informatique pour la paramétrisation des délais dans le mécanisme d'horloge circadienne chez les mammifères  
Porteur : Olivier F. Roux (IRCCyN, École Centrale de Nantes)

Démarré en novembre 2011

Durée : 1 an

- Projet UNIT sur l'apprentissage social (budget : 40.000€)

Objet : Expérimentation d'usage d'un réseau social pour favoriser l'apprentissage

Porteur : Christian Colin (École des Mines de Nantes)

Démarré en janvier 2012

Durée : 1 an

- Projet UNIT TCS (budget : 40.000€)

Objet : Extension des tests de concordance de scripts pour l'évaluation des compétences en informatique

Porteur : Christian Colin (École des Mines de Nantes)

Démarré en mars 2008

Durée : 2 ans

### 1.2.5.3 Autres

Je suis titulaire de la Prime d'Excellence Scientifique (PES) depuis 2012.

## 1.2.6 Valorisation logicielle

Convaincu de l'importance des implantations pratiques en informatique, j'incite mes étudiants à valoriser leur travail dans des logiciels distribués sous licence libre [114]. C'est ainsi que les résultats obtenus au cours des thèses de Loïc Paulevé et Maxime Folschette ont été implantés dans Pint (<http://loicpauleve.name/pint/>), logiciel que Loïc maintient et anime. J'ai par ailleurs pris part à la supervision d'un projet visant à doter Pint d'une interface graphique : gPH (<http://loicpauleve.name/pint/gui/>).

Par ailleurs, je co-supervise avec Karen Reid (Université de Toronto) le développement du logiciel libre MarkUs, pour l'évaluation des travaux d'étudiants en travaux pratiques d'informatique. Cette collaboration a donné lieu à un prix (Trophées des Technologies Éducatives au salon Educatec-Educative 2011), une publication en conférence internationale (ESDA [117]), une nationale (TICE [116]) et à des exposés aux Rencontres Mondiales du Logiciel Libre (2009 et 2011).

## 1.2.7 Cours et séminaires invités

### 1.2.7.1 Cours invités de niveau doctoral

J'ai été sollicité pour assurer les cours suivants :

- Série de 4 conférences de 2h sur le thème *Computational Approaches to Analyze Complex Dynamic Systems : Model-Checking and its Applications* au National Institute of Informatics (Tôkyô, Japon), au printemps 2013 (21 mars, 3, 10 et 17 avril 2013), dans le cadre des "NII International advanced lectures series on ICT"<sup>2</sup> ;
- Cours invité de 4h sur le thème *Modélisation du temps pour la vérification des systèmes dynamiques* (cours conjoint avec Damien Eveillard et Olivier F. Roux). École Jeunes Chercheurs en Informatique Mathématique 2012 (GDR Informatique Mathématique), Rennes, France, Mars 2012.

### 1.2.7.2 Séminaires invités

J'ai été sollicité pour effectuer les séminaires suivants :

- *Towards the combination of LF1T and model-checking*. Invited talk at Workshop on Learning from Interpretation Transition (LFIT), Wakayama University, Wakayama, Japan, June 2015 (Tue. 2015/06/23).
- *Parametric model-checking to assess resilience properties of oscillatory biological systems*. Invited talk at Workshop for Computational Techniques for Sustainability and Resilience in Ishigaki, Japan, March 2015 (Sun. 2015/03/01).
- *Benefits of parametric model-checking to assess the resilience of mammalian circadian rhythm*. Invited talk at NII Shonan Meeting on Systems Resilience, Japan, February 2015 (Tue. 2015/02/23).
- *Analyzing the resilience of mammalian circadian rhythm through parametric model-checking*. Invited talk at The 7th International Workshop on Information Systems for Social Innovation 2015, Tôkyô, Japan, February 2015 (Tue. 2015/02/16).
- *Analyzing regulation thanks to a concurrent dynamic framework : the Process Hitting*. Invited talk at Kyoto University Bioinformatics Center, Kyoto, Japan, February 2015 (Fri. 2015/02/06).
- *Analyzing the dynamics of concurrent systems thanks to model-checking based approaches*. Invited talk at Center for Information and Neural Networks (CiNet), NICT, Osaka, Japan, January 2015 (Fri. 2015/01/09).
- *Analyzing the Dynamics of Large-Scale Concurrent Systems with Process Hitting*. Invited talk at Meeting on Abductive and Inductive Reasoning, Wakayama University, Wakayama, Japan, December 2014 (Tue. 2014/12/16).
- *Why Model-Checking is Significant for Biology ?* Invited talk at KUBIC-NII Joint Seminar on Bioinformatics 2014, Kyoto University, Japan, November 2014 (Mon. 2014/11/17).

---

2. <http://www.nii.ac.jp/en/about/international/ictlec/>

- *Towards the use of Process Hitting to process biological observations inconsistent with background knowledge* (joint talk with Maxime Folschette and Yoshitaka Yamamoto). Invited talk at the NII-Yamanashi-LRI Workshop, LRI, Orsay, France, October 2014 (Mon. 2014/10/06).
- *Resilience in the mammalian circadian clock*. Invited talk at Systems Resilience Workshop, Karuizawa, Japan, September 2014 (Mon. 2014/09/01).
- *Challenges of model-checking of hybrid systems*. Invited Talk at 5th CSPSAT2 meeting in Kobe University, Rokko campus, Japan, August 2014 (Thu. 2014/08/21).
- *Analyzing the Dynamics of Large Biological Regulatory Networks with Process Hitting*. Invited talk at Yamanashi University, Kofu, Japan, June 2014 (Wed. 2014/06/18).
- *Towards the analysis and inference of large biological models* (joint talk with Maxime Folschette). Invited talk at the 5th JFLI-NII-LRI Workshop on Formal Approaches for Modeling and Analyzing Biological Networks, LRI, Orsay, France, October 2013 (Thu. 2013/10/10).
- *Formal methods for the inference of discrete and timing parameters in large-scale dynamic systems*. Invited talk at the Japanese French Laboratory for Informatics (JFLI), Tôkyô, Japan, May 2013 (Thu. 2013/05/09).
- *Inference in Biological Regulatory Networks : From Discrete Parameters to Timing Information*. Invited talk at KUBIC-NII Joint Seminar on Bioinformatics 2013, Kyoto University, Japan, April 2013.
- *Formal Models for the Inference of Time Delay Parameters in Biological Regulatory Networks*. Invited talk at Inoue Laboratory, National Institute of Informatics (NII), Japan, January 2013.
- *Modèles pour l'inférence des paramètres temporels et stochastiques des réseaux de régulation biologiques* (joint work with Loïc Paulevé and Olivier F. Roux). Séminaire à l'I3S, Nice, France, novembre 2009.
- *À la recherche de méthodes efficaces pour la vérification de réseaux de Petri à chronomètres en temps discret* (joint work with Didier Lime, Pierre Molinaro et Olivier H. Roux). Journée AFSEC "Structures de données pour l'analyse des Réseaux de Petri", France, octobre 2008.
- *À la recherche de méthodes efficaces pour la vérification de réseaux de Petri à chronomètres en temps discret* (joint work with Didier Lime, Pierre Molinaro et Olivier H. Roux). Séminaire au LSV, Cachan, France, mai 2008.
- *Méthode symbolique pour le calcul de l'espace d'états des réseaux de Petri à chronomètres en temps discret* (joint work with Didier Lime, Pierre Molinaro et Olivier H. Roux). Séminaire aux journées STP du GDR MACS, Roanne, France, mars 2008.

- *À la recherche de méthodes efficaces pour la vérification de réseaux de Petri à chronomètres en temps discret* (joint work with Didier Lime, Pierre Molinaro et Olivier H. Roux). Séminaire au LISTIC, Annecy, France, mars 2008.
- *Time Petri nets with stopwatches : discrete-time or dense-time ?* (joint work with Didier Lime, Pierre Molinaro et Olivier H. Roux) Séminaire au LaBRI, Bordeaux, France, octobre 2006.

J'ai par ailleurs effectué douzaine d'exposés, dans le cadre de groupes de travail régionaux (séminaire Vivaldi, matinée du e-learning, ...) ou nationaux (Rencontres Mondiales du Logiciel Libre, journée Syntec-Ingénierie, ...) ayant pour thème des retours d'expériences sur des projets pédagogiques innovants. La liste détaillée est disponible sur <http://morganmagnin.net/tice/#publications>

### 1.2.8 Médias

J'ai pris part à différentes actions de vulgarisation dans les médias. Cela s'est traduit par des interviews dans des médias télévisés (Télé Nantes), radiophoniques (Prun', Euradio), papiers (Le Monde, 01 Informatique, Inffo formation) et web (EducPros, Thot Cursus). Une liste (non-exhaustive) de ces interventions peut également être trouvée sur mon site web : <http://morganmagnin.net/tice/#publications>

J'ai également eu l'occasion de participer à des revues institutionnelles :

- J.-P. ELLOY AND M. MAGNIN, *Le numérique : une bibliothèque universelle pour la création scientifique ?*, Revue de la BNU, 10 (2014), pp. 68–75
- M. MAGNIN, *Relever le défi des moocs*, Arabesques, 71 (2013), pp. 20–21

### 1.2.9 E-learning

Dès mon recrutement en tant que maître de conférences à l'École Centrale de Nantes en 2008, j'ai été co-responsable du projet HP "Technology for Teaching" baptisé *Most complex systems are designed with the simplest tools*, que j'avais co-rédigé avec Guillaume Moreau, alors DSI de l'établissement. Cette implication m'a conduit à ensuite être nommé, en 2010, chargé de mission "EAT-TICE" (Enseigner et Apprendre avec les Technologies - Les TICE à Centrale Nantes), placée sous la responsabilité de la Direction de la Formation et dotée d'un budget spécifique (environ 50.000€ par an) pour suivre, questionner et mettre en œuvre les innovations pédagogiques et le recours aux TICE. Cette implication a continué de prendre de l'ampleur suite à différents recrutements menés et la création d'un véritable département TICE & Pédagogie au sein de l'École Centrale de Nantes, département dont je suis responsable (5 personnes, 2.88 ETP).

Dans le cadre de ces missions, j'ai été amené à explorer différents thèmes dans le domaine de la pédagogie numérique. D'abord dans l'utilisation, puis dans la conception, de méthodologies et d'outils innovants. Parce que ces actions ont également contribué à ma prise d'autonomie, ont



donné lieu à différentes publications dans des conférences nationales, internationales ou des revues, et à l'encadrement d'une thèse, je tenais à les présenter dans cette sous-partie.

### 1.2.9.1 Mobilité

De par leur qualité d'élèves-ingénieurs, les étudiants de notre établissement sont régulièrement placés en situation de mobilité (stage en entreprise, projet mené en collaboration avec un partenaire industriel, semestre à l'étranger, etc.). Cette situation motive des projets spécifiques visant tout autant à rendre l'établissement accessible de partout qu'à accompagner les étudiants à utiliser les technologies les plus pertinentes pour y parvenir.

Sur ce plan, nous avons notamment (mais pas exclusivement) développé des actions selon deux axes :

- Usage des technologies tactiles (Tablets PC) pour soutenir l'implication et l'autonomie des étudiants dans leurs études : en septembre 2008, nous avons été lauréats du concours HP Technology for Teaching (100 000\$) pour un projet d'insertion des Tablets PC dans l'enseignement supérieur. Cela nous a permis d'obtenir une vingtaine de machines. Cet effort s'est prolongé en 2012 par l'achat d'un nouveau parc d'une vingtaine de machines. Celles-ci sont utilisées par les enseignants et les étudiants dans des modules spécifiquement conçus pour cela, aussi bien en cours magistraux, TD que TP. En pratique, cette initiative impacte une trentaine d'enseignants par an, et la totalité de chaque promotion à l'issue des trois années passées à l'école. En outre, dans le cadre de leur contrat d'apprentissage, tous les élèves de la filière par alternance de la formation généraliste se voient prêter un Tablet PC pour la durée de leurs études. L'usage de ces machines a contribué à l'émergence d'un esprit de groupe. Les leçons de ces différentes expériences ont été synthétisées notamment dans [112].
- Expérimentations autour des réseaux sociaux (blogs, Twitter, etc.) : dans le contexte de certaines filières, un blog a été mis en place pour permettre un lien plus facile entre étudiants, enseignants et partenaires industriels. Le blog sert également d'outil de partage de veille documentaire, non seulement à l'intérieur de l'établissement, mais également vis-à-vis de l'extérieur [60]. De plus, après plusieurs expérimentations d'usage de Twitter dans un contexte d'enseignement et d'apprentissage, nous avons pris part à un projet UNIT sur l'émergence des réseaux sociaux dans l'apprentissage.
- Médias enrichis pour un apprentissage en autonomie : au cours de ces dernières années, les nouveaux formats de livres électroniques ont fait l'objet de recherches dédiées. Entre autres, l'établissement a vu naître un objet novateur issu de plusieurs de ses laboratoires : eZoomBook [59]. Avec une approche complémentaire à ces travaux, nous avons entamé une étude visant à explorer l'apport possible, dans une démarche d'apprentissage (guidé ou non), d'ontologies de domaine construites à partir de livres de référence. Notre champ

de recherche se focalise pour le moment sur les sciences de l'ingénieur et fait l'objet de la thèse de Simon Carolan [25, 26].

### 1.2.9.2 Évaluation

L'apprentissage par autonomie requiert l'accès non seulement à des ressources documentaires mais également à des outils d'auto-évaluation. C'est pour répondre à ce besoin que la fourniture de QCM s'est déjà montrée utile pour permettre aux étudiants de tester leurs connaissances. C'est dans cet objectif que nous avons pris part à un projet UNIT sur les Tests de Concordance de Scripts (TCS). Si les QCM servent avant tout à évaluer des connaissances, les TCS ont pour objectif d'évaluer le raisonnement d'étudiants placés dans une situation professionnelle. Initialement, les TCS s'appliquaient à des situations incertaines du milieu médical. Leur principe nous est apparu très similaire à des cas pratiques de diagnostic en ingénierie, d'où la motivation à les étendre aux sciences de l'ingénieur. Les TCS sont actuellement utilisés dans le cadre de l'évaluation des compétences en informatique [113, 115] et nous avons pour projet de les mettre en œuvre dans d'autres disciplines.

D'un point de vue outillage, nous avons participé au développement et déploiement de MarkUs<sup>3</sup>, application web permettant de rassembler et corriger directement les travaux remis par les étudiants dans le cadre de TP ou projets d'informatique. Les bénéfices et limites d'un tel outil dans l'apprentissage ont été présentés dans [114, 116, 117].

### 1.2.9.3 MOOC

En 2012, j'ai participé à la mise en place du premier MOOC (*Massive Online Open Course*) francophone, ITyPA ("Internet Tout y est Pour Apprendre") d'octobre à décembre 2012 [75]. Chaque session de ce MOOC s'étale sur 10 semaines. Ce module est axé autour de la question : "comment apprendre en ligne?". Il balaie un large panel de techniques, outils et méthodes pour repérer les sources d'information les plus intéressantes sur la toile, et contribuer à son tour à l'augmentation du savoir partagé dans un ou plusieurs domaines. Plus de 1300 personnes ont participé à cette première édition.

À l'automne 2013, nous avons proposé une nouvelle itération d'ITyPA en approfondissant la question des certificats dans un MOOC. Les participants ont eu le choix d'assister librement aux séances ou de s'investir davantage dans le but d'obtenir non pas un simple certificat mais un badge délivré à partir d'une évaluation personnelle. Celle-ci est alignée sur le référentiel de la Fondation Mozilla dans le cadre l'Open Badges Initiative qui permet la reconnaissance de compétences acquises sur le web ou en présentiel. Plus de 2000 personnes ont pris part à cette seconde édition.

Ces expériences ont été riches d'enseignements [27, 29] pour la mise en place ultérieure de plusieurs MOOCs [30].

---

3. <http://markusproject.org/>

#### 1.2.9.4 Collaborations

Les travaux dans le domaine des innovations pédagogiques et des TICE font appel à de nombreuses collaborations :

- avec la Conférence Régionale des Pays de Loire et la COMUE UBL (autour de la formation des enseignants et de la balado-diffusion) ;
- avec les écoles du Groupe des Écoles Centrales (autour des technologies mobiles pour l'enseignement) ;
- avec l'UNT UNIT (autour des TCS et des réseaux sociaux) ;
- avec l'Université de Toronto (sur le logiciel libre MarkUs) ;
- avec Hewlett-Packard (autour de l'utilisation de Tablets PC en pédagogie).

### 1.3 Publications

#### 1.3.1 Revues internationales avec comité de lecture

- (R1) N. SCHWIND, M. MAGNIN, L. INOUE, T. OKIMOTO, T. SATO, K. MINAMI, AND H. MARYAMA, *Formalization of resilience for constraint-based dynamic systems*, Journal of Reliable Intelligent Environments, (2015), pp. 1–19
- (R2) M. FOLSCHETTE, L. PAULEVÉ, M. MAGNIN, AND O. ROUX, *Sufficient conditions for reachability in automata networks with priorities*, Theoretical Computer Science, 608 (2015), pp. 66–83
- (R3) F. CHINESTA, M. MAGNIN, O. ROUX, A. AMMAR, AND E. CUETO, *Kinetic theory modeling and efficient numerical simulation of gene regulatory networks based on qualitative descriptions*, Entropy, 17 (2015), pp. 1896–1915
- (R4) T. RIBEIRO, M. MAGNIN, K. INOUE, AND C. SAKAMA, *Learning delayed influences of biological systems*, Frontiers in Bioengineering and Biotechnology, 2 (2015), p. 81
- (R5) M. FOLSCHETTE, L. PAULEVÉ, K. INOUE, M. MAGNIN, AND O. ROUX, *Identification of biological regulatory networks from process hitting models*, Theoretical Computer Science, 568 (2015), pp. 49–71
- (R6) L. PAULEVÉ, M. MAGNIN, AND O. ROUX, *Static analysis of biological regulatory networks dynamics using abstract interpretation*, Mathematical Structures in Computer Science, 22 (2012), pp. 651–685
- (R7) L. PAULEVÉ, M. MAGNIN, AND O. ROUX, *Tuning temporal features within the stochastic  $\pi$ -calculus*, Software Engineering, IEEE Transactions on, 37 (2011), pp. 858–871
- (R8) L. PAULEVÉ, M. MAGNIN, AND O. ROUX, *Refining dynamics of gene regulatory networks in a stochastic  $\pi$ -calculus framework*, in Transactions on computational systems biology XIII, Springer, 2011, pp. 171–191

- (R9) M. MAGNIN, P. MOLINARO, AND O. H. ROUX, *Expressiveness of petri nets with stopwatches. discrete-time part*, Fundamenta Informaticae, 97 (2009), pp. 139–176
- (R10) M. MAGNIN, P. MOLINARO, AND O. H. ROUX, *Expressiveness of petri nets with stopwatches. dense-time part*, Fundamenta Informaticae, 97 (2009), pp. 111–138

### 1.3.2 Revues nationales avec comité de lecture

- (R11) L. PAULEVÉ, M. FOLSCHETTE, M. MAGNIN, AND O. ROUX, *Analyses statiques de la dynamique des réseaux d'automates indéterministes*, Revue des Sciences et Technologies de l'Information - Série TSI : Technique et Science Informatiques, 34 (2015), p. 22
- (R12) M. MAGNIN AND G. MOREAU, *Utilisation des tests de concordance de scripts pour l'évaluation en informatique*, Revue Sticef.org, 9 (2011), p. 01

### 1.3.3 Chapitres de livres

- (L1) L. PAULEVÉ, C. CHANCELLOR, M. FOLSCHETTE, M. MAGNIN, AND O. ROUX, *Logical Modeling of Biological Systems*, Wiley, 2014, ch. Analyzing Large Network Dynamics with Process Hitting, pp. 125 – 166
- (L2) G. MOREAU, M. SERVIÈRES, J.-M. NORMAND, AND M. MAGNIN, *Challenges of image-based crowd-sourcing for situation awareness in disaster management*, in Improving Disaster Resilience and Mitigation-IT Means and Tools, Springer, 2014, pp. 103–118
- (L3) A. DAVID, G. BEHRMANN, P. BULYCHEV, J. BYG, T. CHATAIN, K. LARSEN, P. PETTERSSON, J. RASMUSSEN, J. SRBA, W. YI, K. JOERGENSEN, D. LIME, M. MAGNIN, O. ROUX, AND L.-M. TRAONOUÉZ, *Communicating Embedded Systems - Software and Design*, ISTE Publishing/John Wiley, 2009, ch. Tools for Model-Checking Timed Systems
- (L4) A. DAVID, G. BEHRMANN, K. LARSEN, P. PETTERSSON, J. RASMUSSEN, W. YI, D. LIME, M. MAGNIN, AND O. ROUX, *Approches formelles des systèmes embarqués communicants*, Hermes Lavoisier, 2008, ch. Outils de Model Checking

### 1.3.4 Conférences internationales avec comité de lecture et actes

Les conférences CMSB, ICATPN et CAV ont des taux d'acceptation inférieurs à 45%. Les 5 publications en conférences les plus significatives sont : (C1, C7, C10, C12, C13) qui ont été publiées dans les *Lecture Notes in Computer Science* (respectivement : 9308, 8130, 7605, 5062 et 3576).

- (C1) E. BEN ABDALLAH, M. FOLSCHETTE, O. ROUX, AND M. MAGNIN, *Exhaustive analysis of dynamical properties of biological regulatory networks with answer set programming*, in 2015 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), IEEE, 2015, pp. 281–285

- (C2) A. ANDREYCHENKO, M. MAGNIN, AND K. INOUE, *Modeling of resilience properties in oscillatory biological systems using parametric time petri nets*, in 13th International Conference on Computational Methods in Systems Biology - , CMSB 2015, Nantes, France, O. F. Roux and J. Bourdon, eds., vol. 9308 of Lecture Notes in Computer Science, Springer, 2015
- (C3) S. CAROLAN, M. MAGNIN, AND A.-L. KABALU, *Sparkling a digital revolution: Digital educational tools in fragile and emerging learning contexts*, in Digital Intelligence 2014, 2014
- (C4) S. CAROLAN, F. CHINESTA, C. EVAIN, M. MAGNIN, AND G. MOREAU, *Extracting domain ontologies from reference books*, in IEEE 14th International Conference on Advanced Learning Technologies (ICALT), IEEE, 2014, pp. 544–545
- (C5) T. RIBEIRO, M. MAGNIN, AND K. INOUE, *Learning delayed influence of dynamical systems from interpretation transition*, in The 24th International Conference on Inductive Logic Programming, 2014.  
Short paper (<http://tony.research.free.fr/paper/ILP2014short>)
- (C6) S. CAROLAN, M. MAGNIN, AND J.-M. GILLIOT, *Engineering moocs for future engineers: Integrating moocs into formal learning environments*, EMOOCs 2014: 2nd MOOC European stakeholders summit, (2014)
- (C7) S. CAROLAN, F. CHINESTA, C. EVAIN, M. MAGNIN, AND G. MOREAU, *Towards augmented learning in science and engineering in higher education*, in IEEE 13th International Conference on Advanced Learning Technologies (ICALT), IEEE, 2013, pp. 512–513
- (C8) C. CHANCELLOR, A. AMMAR, F. CHINESTA, M. MAGNIN, AND O. ROUX, *Linking discrete and stochastic models: The chemical master equation as a bridge between process hitting and proper generalized decomposition*, in Computational Methods in Systems Biology, A. Gupta and T. Henzinger, eds., vol. 8130 of Lecture Notes in Computer Science, Springer, 2013, pp. 50–63
- (C9) M. MAGNIN, G. MOREAU, N. VAROQUAUX, B. VIALLE, K. REID, M. CONLEY, AND S. GEHWOLF, *Markus: An open-source web application to annotate student papers on-line*, in ASME 2012 11th Biennial Conference on Engineering Systems Design and Analysis, American Society of Mechanical Engineers, 2012, pp. 301–307
- (C10) C. EVAIN, S. CAROLAN, AND M. MAGNIN, *Promoting science and engineering to generation z*, in ASME 2012 11th Biennial Conference on Engineering Systems Design and Analysis, American Society of Mechanical Engineers, 2012
- (C11) M. FOLSCHETTE, L. PAULEVÉ, K. INOUE, M. MAGNIN, AND O. ROUX, *Concretizing the process hitting into biological regulatory networks*, in Computational Methods in Systems Biology, D. Gilbert and M. Heiner, eds., Lecture Notes in Computer Science, Springer, 2012, pp. 166–186

- (C12) M. MAURIN, M. MAGNIN, AND O. ROUX, *Modeling of genetic regulatory network in stochastic  $\pi$ -calculus*, in Bioinformatics and Computational Biology, Springer, 2009, pp. 282–294
- (C13) M. MAGNIN, D. LIME, AND O. H. ROUX, *Symbolic state space of stopwatch petri nets with discrete-time semantics*, in The 29th International Conference on Application and Theory of Petri Nets and other models of concurrency (ICATPN 2008), K. van Hee and R. Valk, eds., vol. 5062 of Lecture Notes in Computer Science, Xi'an, China, June 2008, Springer, pp. 307–326
- (C14) G. GARDEY, D. LIME, M. MAGNIN, AND O. ROUX, *Romeo: A tool for analyzing time petri nets*, in Computer Aided Verification, K. Etessami and S. Rajamani, eds., vol. 3576 of Lecture Notes in Computer Science, Springer, 2005, pp. 418–423

### 1.3.5 Conférences nationales avec comité de lecture et actes

Les conférences QPES et TICE font partie des colloques francophones majeurs dans le domaine du e-learning. Les taux d'acceptation de QPES et des axes "retours d'expérience" et "communications scientifiques" de TICE sont généralement 75%, 70% et 25% respectivement.

- (C15) S. CAROLAN, M. MAGNIN, J.-M. GILLIOT, A.-C. GROLLEAU, AND C. VAUFREY, *Un mooc et ses partenaires*, in 9ème Colloque International Technologies de l'Information et de la Communication pour l'Enseignement (TICE 2014). Retour d'Expérience (ReX), 2014
- (C16) N. SCHWIND, M. MAGNIN, AND K. INOUE, *Resilience of event-driven dynamic systems*, in 27th Annual Conference on Japanese Society of Artificial Intelligence (JSAI 2013), vol. 27, 2013, pp. 1–8
- (C17) S. CAROLAN AND M. MAGNIN, *Le mooc et la motivation : les élèves face à une formation autogérée*, in 6ème Conférence Environnements Informatiques pour l'Apprentissage Humain (EIAH 2013), Atelier MOOC, 2013
- (C18) J.-M. GILLIOT, A.-C. GROLLEAU, M. MAGNIN, AND C. VAUFREY, *Itypa, un premier mooc francophone et connectiviste*, in QPES 2013: colloque questions de pédagogies dans l'enseignement supérieur, 2013
- (C19) M. MAGNIN, G. MOREAU, V. TOURRE, N. VAROQUAUX, AND B. VIALLE, *Apports de l'application web libre markus dans le processus de correction de travaux en informatique*, in 8ème Colloque International Technologies de l'Information et de la Communication pour l'Enseignement (TICE 2012). Retour d'Expérience (ReX), 2012
- (C20) M. MAGNIN, *Les interfaces entre le manga et la société japonaise : renforcer l'utopie pour affronter la crise*, in 4èmes Rencontres Internationales Jules Verne, 2012
- (C21) M. MAGNIN AND G. MOREAU, *La professionnalisation des étudiants en informatique par les projets libres*, in 6eme Colloque Questions de pédagogies dans l'enseignement supérieur (QPES), 2011

- (C22) C. EVAIN, M. MAGNIN, AND G. MOREAU, *Diffuser les bonnes pratiques autour de l'usage des blogs en pédagogie*, in 7ème Colloque International Technologies de l'Information et de la Communication pour l'Enseignement (TICE 2010). Retour d'Expérience (ReX), 2010
- (C23) M. MAGNIN AND G. MOREAU, *Retour sur l'usage de tablet pc à l'école centrale de nantes : autonomie et initiatives*, in 7ème Colloque International Technologies de l'Information et de la Communication pour l'Enseignement (TICE 2010). Retour d'Expérience (ReX), 2010
- (C24) M. MAGNIN AND G. MOREAU, *Utilisation des tests de concordance de scripts pour l'évaluation en informatique*, in 7ème Colloque International Technologies de l'Information et de la Communication pour l'Enseignement (TICE'2010). Communication scientifique (ComSci), 2010
- (C25) C. SEIDNER, J. LERAT, O. ROUX, AND M. MAGNIN, *Vérification dynamique et formelle d'un système décrit par son architecture fonctionnelle à l'aide de réseaux de petri temporels (tpn): promesses et perspectives*, in 4e conférence annuelle d'ingénierie système (AFIS), 2006

### 1.3.6 Workshops internationaux avec comité de lecture et actes

Les publications en workshops (W1,W3,W5) ont été publiées dans les *Electronic Notes in Theoretical Computer Science*.

- (W1) M. FOLSCHETTE, L. PAULEVÉ, M. MAGNIN, AND O. ROUX, *Under-approximation of reachability in multivalued asynchronous networks*, *Electronic Notes in Theoretical Computer Science*, 299 (2013), pp. 33–51.  
 Proceedings of the fourth International Workshop on Interactions between Computer Science and Biology (CS2Bio'13)
- (W2) M. FOLSCHETTE, L. PAULEVÉ, K. INOUE, M. MAGNIN, AND O. ROUX, *Abducting biological regulatory networks from process hitting models*, in ECML-PKDD 2012 Workshop on Learning and Discovery in Symbolic Systems Biology, 2012, pp. 24–35
- (W3) L. PAULEVÉ, M. MAGNIN, AND O. ROUX, *Abstract Interpretation of Dynamics of Biological Regulatory Networks*, *Electronic Notes in Theoretical Computer Science*, 272 (2011), pp. 43–56.  
 Proceedings of The First International Workshop on Static Analysis and Systems Biology (SASB 2010)
- (W4) M. MAGNIN, P. MOLINARO, AND O. H. ROUX, *Decidability, expressivity and state-space computation of stopwatch petri nets with discrete-time semantics*, in 8th International Workshop on Discrete Event Systems, IEEE, 2006, pp. 33–38
- (W5) M. MAGNIN, D. LIME, AND O. H. ROUX, *An efficient method for computing exact state space of Petri nets with stopwatches*, in third International Workshop on Software Model-Checking (SoftMC'05), vol. 144 of *Electronic Notes in Theoretical Computer Science*, Edinburgh, Scotland, UK, July 2005, Elsevier, pp. 59–77

## INTRODUCTION

Ce manuscrit, centré sur ma contribution dans le domaine des méthodes formelles pour la biologie, rend compte des recherches que j'ai menées dans le domaine depuis mon recrutement en tant que maître de conférences à l'École Centrale de Nantes en 2008. Il s'agit là d'une vision générale sur les travaux que j'ai menés, avec mes collègues de l'IRCCyN, les doctorants que j'ai encadrés et d'autres partenaires. Le détail des contributions est disponible dans l'ensemble des publications référencées.

## 2.1 Les défis de la modélisation en biologie : une vision quantitative ou qualitative ?

La régulation est un aspect central - et crucial - des systèmes biologiques, que nous les considérons à l'échelle génétique [11], métabolique [166] ou écologique [132]. Acquérir une compréhension précise des mécanismes de régulation est l'un des principaux objectifs de la biologie des systèmes [92, 93]. Cette discipline, dont l'essor remonte au début des années 2000, résulte de la synergie entre la biologie cellulaire et la cybernétique [171]. Elle suscite les collaborations entre biologistes, physiciens, mathématiciens et informaticiens [83].

Les travaux que je synthétise dans ce manuscrit se situent dans cette lignée. Ils tournent autour de l'étude des interactions des composants d'un système biologique, et de l'impact de ces interactions sur la fonctionnalité et le comportement (dynamique) du système. Comme je le détaillerai, mes recherches portent sur la conception de méthodes formelles visant à une meilleure compréhension du fonctionnement de systèmes biologiques complexes. Cependant, même si les approches auxquelles j'ai contribué portaient essentiellement sur des applications génétiques, leur champ d'application potentiel est bien plus large. En fait, les travaux présentés dans ce



document peuvent servir à l'étude des systèmes réactifs, qu'il s'agisse de réseau de régulation biologique, d'un processus logistique ou d'un système électronique embarqué.

Les progrès récents dans le domaine de la biologie moléculaire ont permis de produire des cartes complètes du génome de nombreux organismes vivants. Seulement, ces cartes statiques - popularisées chez l'humain notamment à travers le "Human Genome Project" [42] - ne suffisent pas à rendre compte de la complexité intrinsèque du vivant. Il est en effet nécessaire de considérer les phénomènes génétiques dans toute la complexité de leurs interactions, avec leur environnement (par exemple, le stress que représente un composant extérieur tel que la lumière ou l'ingestion d'une protéine) et leur dynamique.

Dans le même temps, le développement de biotechnologies telles que les puces à ADN et l'essor de techniques de séquençage à haut-débit (tel le RNAseq) facilitent la production de données de séries temporelles correspondant à l'expression de plusieurs milliers de gènes. Il se pose dès lors la question du sens à donner à cette profusion d'information. L'un des principaux défis actuels de la biologie des systèmes est ainsi d'intégrer ces données à haut débit et d'en inférer les réseaux de régulation génétique associés. Puisque les interactions se produisent à différentes échelles (gènes, protéines, constituants biochimiques, cellules, etc.), il est nécessaire d'élaborer des méthodes qui peuvent formellement et automatiquement apprendre les données biologiques pour faciliter la compréhension, à un niveau systémique, des phénomènes en jeu.

La modélisation des mécanismes de régulation biologique se décompose en deux principales tendances. La première, quantitative, repose sur les équations différentielles ordinaires mettant en jeu l'expression quantitative des composants en interaction. Cependant, ces équations sont généralement non-linéaires, ce qui empêche leur résolution analytique. De plus, les données biologiques obtenues expérimentalement sont généralement relativement bruitées, ce qui impose de les filtrer efficacement. A contrario, la deuxième tendance consiste à traiter le problème via une méthodologie discrète. Autrement dit, l'expression de chaque composant est évaluée selon plusieurs (généralement deux, mais des extensions existent, comme nous le verrons dans la section suivante) niveaux qualitatifs, et non plus quantitativement. Même si une telle modélisation discrète pourrait être considérée comme une abstraction moins fidèle de la réalité, elle s'est révélée efficace dans le traitement de nombreuses questions biologiques qualitatives (telles que la compréhension des interactions entre plusieurs composants d'un système biologique, ou la détermination de l'accessibilité d'un état).

Dans ce manuscrit, nous nous focalisons sur les processus de régulation à l'échelle génétique. Nous choisissons ici d'abstraire certains mécanismes en jeu, par exemple à l'échelle moléculaire (qui requiert de spécifier de manière détaillée le comportement de chaque élément biologique à l'instar des protéines, plasmides, etc.). Il est toutefois bon de noter que plusieurs travaux s'attaquant spécifiquement à ces problématiques ont vu le jour au cours de la dernière décennie, à travers le langage  $\kappa$  [49], le  $\pi$ -calcul stochastique [98, 99], Bio-PEPA [40] ou encore DSD [146].

## 2.2 Le contexte historique : modélisation qualitative booléenne ou multi-valuée

Du fait du bruit expérimental, les données exploitées dans un contexte biologique sont souvent plus qualitatives que quantitatives. C'est de fait assez naturellement qu'une autre démarche d'analyse a émergé à la fin des années 60 à travers la modélisation qualitative. Il est significatif que les fondements d'une telle démarche aient été posés, en parallèle et indépendamment l'un de l'autre : d'une part par Stuart Kauffman via les réseaux booléens [90], d'autre part par René Thomas dans les réseaux auxquels il a donné son nom [168]. Ces deux formalismes abstraient les gènes (plus exactement : les protéines produites par les gènes) sous la forme de variables booléennes. Ces variables, ou nœuds, ne peuvent prendre que deux états : activé ou désactivé. Les variables booléennes sont liées les unes aux autres par des relations d'activation et d'inhibition, respectivement représentées par des arcs positifs et négatifs. Les différences entre les modèles de Stuart Kauffman et de René Thomas résident notamment dans la fonction de mise à jour, synchrone dans le premier cas, asynchrone dans le second. Nous discutons plus précisément ces distinctions dans les paragraphes suivants.

Tandis que Stuart Kauffman a établi une représentation formelle des interactions entre les composants d'un réseau de régulation génétique par l'intermédiaire d'une modélisation idéalisée, René Thomas s'est appuyé sur les mécanismes régissant la régulation de la transcription pour en proposer une description logique globale. De cette différence de point de vue résultent deux modélisations certes différentes, mais finalement proches dans l'esprit. Dans ses réseaux booléens, Stuart Kauffman considère l'état des gènes comme étant décrit par des variables booléennes, le prochain état résultant de l'application d'une fonction booléenne synchrone régissant l'état de chaque variable. René Thomas, quant à lui, examine l'influence d'une activation (resp. une inhibition) sur un gène cible en fonction d'une valeur seuil : quand le niveau associé au gène en amont de la régulation est supérieur ou égal au seuil, le gène a un effet positif (resp. négatif) ; quand le niveau associé au gène en amont de la régulation est inférieur au seuil, alors le gène a un effet négatif (resp. positif) [154].

Comme on le voit, même si les spécifications de ces deux modèles sont différentes, ils s'appuient tous les deux sur l'idée centrale que l'état suivant dépend de l'état courant via une fonction booléenne. La principale différence entre les deux paradigmes se trouve dans la fonction de mise à jour : alors que les réseaux booléens sont principalement étudiés dans leur sémantique synchrone, les réseaux de René Thomas sont classiquement équipés d'une sémantique asynchrone (correspondant au fait qu'il est très improbable que deux composants biologiques changent simultanément d'état). Des travaux récents [9, 130, 131] ont discuté les différences, pour la dynamique du réseau, d'une sémantique synchrone par rapport à une sémantique asynchrone.

Ces deux paradigmes de modélisation ont été considérés dignes d'intérêt et ont été suivis de nombreux travaux après leur publication. Par exemple, sur les réseaux booléens, les auteurs de [3]

passent en revue un panel d’algorithmes pour traiter l’inférence, l’identification des attracteurs et le contrôle des réseaux booléens. Toutefois les approches classiques échouent généralement à traiter des modèles de grande taille parce que l’espace d’états croît exponentiellement avec le nombre de composants. De fait, un nombre important de travaux ont eu pour objectif d’explicitier la relation entre le graphe d’interaction d’un réseau (c’est-à-dire la donnée des nœuds et des interactions, activatrices ou inhibitrices, entre les nœuds<sup>1</sup>) et ses propriétés dynamiques. Par exemple, certains travaux novateurs se concentrent sur les boucles de rétroaction, i.e., les chaînes circulaires d’interactions. Il est en effet possible de montrer que le nombre d’inhibitions dans une boucle permet d’exprimer une condition nécessaire à l’existence de plusieurs points fixes. Dans [143], les auteurs ont proposé un bilan sur les relations entre le graphe d’interaction et les propriétés dynamiques des réseaux booléens.

Récemment, d’autres approches ont également émergé. L’une d’elle repose sur la conception de méthodes de réduction qui préservent les propriétés dynamiques essentielles du système [18].

Un autre angle d’attaque consiste à analyser les réseaux booléens sous la forme de programmes logiques. Dans [84, 87], les auteurs ont ainsi exhibé une traduction directe des réseaux booléens en programmes logiques et ont proposé des méthodes exploitant ces programmes logiques pour déterminer les trajectoires et les attracteurs de ces réseaux booléens.

Ces paradigmes booléens peuvent paraître comme des modèles simplifiés, mais ils ont conduit à des résultats significatifs sur le comportement des réseaux, tels que la présence de cycles ou d’états stables. De plus, ces modèles ont été étendus au fil des ans pour, par exemple, tenir compte de niveaux supplémentaires d’expressions, ouvrant la voie à des réseaux discrets tels que nous les considérerons dans la suite de ce manuscrit. Ces réseaux discrets permettent d’affiner l’étude des comportements pour distinguer, par exemple, certains types de phénomènes oscillatoires (oscillations amorties ou non), des mécanismes de bifurcation, etc.

## 2.3 Les défis de l’analyse

Grâce aux approches qualitatives, il devient plus abordable — bien que cela reste encore très compliqué — d’avoir une compréhension en profondeur des interactions impliquées dans un réseau génétique. Cette méthodologie vise à expliquer comment les composants d’un réseau de régulation génétique sont contrôlés les uns les autres, ou à prédire un comportement sur la base des observations précédentes. Évidemment, comme l’évoque Grégory Batt dans son manuscrit d’HDR [15], le pouvoir prédictif des modèles est, jusqu’à maintenant, relativement limité. Ainsi un modèle construit dans un certain contexte s’étend rarement à des conditions expérimentales légèrement différentes. Pour répondre à cette question, des travaux récents [169] ont visé à

---

1. Notons que la donnée des nœuds et des interactions n’est pas suffisante pour décrire de manière non-ambiguë la dynamique du réseau : en effet, lorsqu’une activation et une inhibition ciblent simultanément un composant, il est nécessaire d’avoir une information supplémentaire sur la force respective de l’activation par rapport à l’inhibition. Cette donnée est capturée soit par la donnée complète de la fonction booléenne de mise à jour dans les réseaux booléens de Kauffman, soit par la table de paramétrisation dans les réseaux de Thomas.

définir une approche de contrôle de la prédictibilité d'un modèle. Les auteurs se sont notamment intéressés à des questions telles que la manière dont les performances prédictives d'un modèle se dégradent lorsque les temps d'observations sont plus distants dans le temps.

Cet enjeu de prédictibilité est d'une importance cruciale, en particulier dans le domaine de la biologie de synthèse et de la conception de médicaments. Les difficultés rencontrées dans l'analyse et la prédiction peuvent être considérées comme relevant de l'une des quatre grandes catégories suivantes :

- L'identification de paramètres : les modèles discrets tels que les réseaux booléens ou les réseaux de Thomas nécessitent non seulement des informations sur la topologie du réseau et le type d'influences — activation ou inhibition — entre les gènes, mais aussi sur la force respective de chaque interaction. Ce type d'information est nécessaire pour déterminer, par exemple, l'influence prédominante lorsqu'un gène est la cible unique de deux influences aux effets opposés. Comme nous l'avons évoqué précédemment, une telle information est représentée soit au cœur d'une fonction de mise à jour des nœuds du réseau, soit au cœur d'une paramétrisation dédiée. Dans un cas comme dans l'autre, ces informations peuvent se déduire des observations biologiques expérimentales par des méthodes informatiques telles que le model-checking ou la programmation par contraintes.
- L'inférence du modèle : plus large que l'identification de paramètres, le problème de l'inférence englobe aussi la détermination de la structure globale du réseau. Compte tenu de données de séries temporelles pouvant être converties (une fois discrétisées) en système de transitions d'états, il peut être important - pour faciliter l'analyse des données - de construire un modèle discret compatible avec les données disponibles.
- L'identification des états stables et des attracteurs, propriétés clés de la plupart des systèmes biologiques.
- Le contrôle du modèle : la capacité à contrôler un modèle de manière à ce qu'il présente un comportement désiré ou, au contraire, à ce que certains comportements indésirables soient évités est étroitement liée à l'analyse des propriétés clés sus-citées. Il s'agit là du plus récent, mais aussi du plus difficile, thème de recherche dans le domaine de la biologie des systèmes, étroitement lié à la problématique de thérapie génique.

Ces difficultés sont des points d'attaque cruciaux pour notre démarche de modélisation et d'analyse. Toutefois, bien que les données disponibles sur les interactions entre les gènes soient de plus en plus conséquentes, il subsiste un manque de données cinétiques quantitatives précises. Comme mentionné précédemment, les données obtenues expérimentalement sont généralement bruitées, ce qui rend difficile leur exploitation afin d'en déduire des valeurs de paramètres. En conséquence, l'identification des paramètres nécessite un raisonnement indirect, qui devient ardu quand le modèle met en jeu plus d'une dizaine d'éléments en interaction ; en raison de l'explosion combinatoire de l'espace d'états, il est extrêmement difficile de gérer de grands réseaux réalistes de régulation.

Comme nous venons de le constater, il est primordial de concevoir des méthodes efficaces pour inférer et analyser des réseaux de régulation. Cette perspective a suscité de nombreux travaux au cours de la décennie écoulée, que ce soit en adoptant différents types de méthodologie, dont le model-checking [16, 21, 53], la programmation logique [77] ou la théorie des jeux [36] pour n'en citer que trois.

## 2.4 Nos objectifs

Ces problématiques ont conduit à l'émergence de ce qu'on appelle désormais la biologie exécutable, dont le but est de fournir des méthodes formelles pour synthétiser automatiquement des modèles à partir d'expériences. Ces approches considèrent en entrée les données d'expression de gènes et en déduisent les régulations connexes. Une limite commune à nombre de méthodologies de ce type réside dans le fait qu'elles ne traitent pas toutes les informations contenues dans les données de séries temporelles considérées en entrée. Par exemple, les informations quantitatives de synchronisation sont abstraites, alors que c'est une caractéristique majeure pour comprendre le comportement dynamique de nombreux modèles.

Même si la plupart des approches existantes ne concernent que des données statiques, on dénote un intérêt croissant pour les algorithmes d'inférence qui intègrent *les aspects temporels*.

Étant donné le portrait global dressé dans ce chapitre, notre objectif porte sur l'élaboration (si possible automatique) de modèles qualitatifs de réseaux de régulation, via des démarches de model-checking (pour l'analyse) et de programmation logique (pour l'inférence). Les spécificités de notre travail résident dans notre attachement à l'étude de ce problème vis-à-vis de la dynamique des systèmes dans un *contexte large-échelle*, autrement dit avec potentiellement plusieurs centaines de composants interagissant.

Le caractère distinctif de notre démarche réside dans la volonté de prendre en compte la dimension temporelle quantitative (au sens du délai entre plusieurs événements) de ces interactions, et ce, dans le cadre de systèmes de grande dimension. Il s'agit là d'un objectif à long terme, qui passe par plusieurs étapes intermédiaires, à savoir la définition d'un formalisme et de méthodes efficaces pour l'analyse de réseaux de grande taille, puis l'enrichissement progressif de ce formalisme et des algorithmes associés.

## 2.5 Organisation du document

Depuis la fin de ma thèse, en 2007, mes travaux ont tenté d'apporter une contribution aux problèmes sus-cités. Ces recherches ont été menées dans le cadre de thèses que j'ai co-supervisées et de collaborations que j'ai établies. Ce manuscrit en présente une synthèse.

Devant l'enjeu que représente la modélisation et l'analyse de la dynamique de réseaux de régulation à grande échelle, l'objectif de la thèse de Loïc Paulevé (2008-2011) consistait à identifier une classe pertinente de modèles formels sur laquelle il serait possible de définir des méthodes

efficaces d'étude des propriétés dynamiques. Le formalisme qui a été introduit à cette occasion, les Frappes de Processus, se sont révélées particulièrement pertinentes dans le passage à l'échelle de certaines analyses. Après avoir discuté les différents critères de modélisation à prendre en compte en biologie, nous introduirons le formalisme des Frappes de Processus dans le chapitre 3. Nous présenterons également brièvement leurs extensions, mises au point récemment au cours de la thèse de Maxime Folschette (2011-2014), afin d'intégrer des mécanismes de préemption (et, de fait, de permettre la représentation d'un temps quantitatif) dans la modélisation. Dans le chapitre 4, nous revenons sur les méthodes d'analyse mises au point sur les Frappes de Processus, leurs mérites et leurs limites. Nous en profitons également pour évoquer les incursions que nous avons opérées dans le champ du stochastique et du continu. Par suite, le chapitre 5 propose des résultats plus récents, consécutifs à l'établissement d'une collaboration solide dans le domaine de l'apprentissage automatique et de la programmation logique avec l'équipe de Katsumi Inoue au National Institute of Informatics (Tôkyô, Japon). Après des travaux portant sur l'identification des paramétrisations associées aux réseaux de régulation, nous nous intéressons de manière plus précise à l'inférence automatique de réseaux de régulation à partir de données de séries temporelles. Ces recherches, effectuées avec Katsumi Inoue et Tony Ribeiro (doctorant avec lequel j'ai fortement collaboré à partir de sa deuxième année de thèse), ont ouvert de nouvelles perspectives, notamment dans le domaine de la complétion de réseaux, thème au cœur de la thèse d'Emna Ben Abdallah (2014-2017). Enfin, le chapitre 6 propose un bilan de mes différentes contributions et résume les pistes de recherche que je souhaite explorer au cours des prochaines années.



## MODÉLISER ET ABSTRAIRE

Ce chapitre introduit la problématique de modélisation et d'abstraction, d'abord dans le contexte le plus général possible, puis en nous penchant plus spécifiquement sur l'apport de la modélisation pour comprendre des mécanismes de régulation. Nous détaillons les choix qui s'offrent au modélisateur, notamment en matière de sémantique. Nous balayons les principaux modèles existants pour présenter ensuite un formalisme, les Frappes de Processus, dont nous avons mis en évidence les avantages pour traiter des modèles de grande taille. Ce modèle, constituant une restriction spécifique des réseaux d'automates asynchrones, permet de définir des méthodes d'analyse efficaces dont nous introduirons les principes dans le chapitre suivant.

Ces travaux ont été initiés dans le cadre de la thèse de doctorat de Loïc Paulevé [133], puis prolongés par la thèse de Maxime Folschette [64].

La définition initiale des Frappes de Processus a fait l'objet d'un article en revue [139]. La présentation du formalisme a ensuite été revue et affinée pour un chapitre de livre [135]. L'enrichissement du formalisme avec des priorités a été introduit dans une publication en workshop [68] étendue ensuite en revue [69].

### 3.1 Pourquoi abstraire ?

Que l'on étudie l'évolution d'un éco-système, d'une fonction biologique chez un humain ou des interconnexions entre programmes informatiques dans un avion, la complexité des interactions en jeu nécessite la constitution de représentations alternatives, plus simples. Celles-ci, généralement complémentaires, ont pour objectif de rendre la complexité d'un système accessible à un esprit humain, ou de la rendre analysable à grande échelle par un processus automatique et rationnel,



c'est-à-dire par un ordinateur.

Depuis quelques années, nous avons changé de paradigme. Nous sommes passés d'une ère où les données étaient difficiles à obtenir, fruits d'expériences coûteuses et/ou compliquées à mener, à une démultiplication de celles-ci. Il en va ainsi de presque tous les domaines scientifiques. L'émergence des objets connectés (par exemple les montres numériques enregistrant des données physiologiques) et des techniques de séquençage à haut-débit en sont deux illustrations parmi d'autres dans le secteur de la santé. Tous domaines confondus, il se produisait ainsi 29 000 GigaOctets (Go) d'information par seconde en 2012 selon une étude IDC sponsorisée par EMC Gartner.

Pour être comprises, analysées, ajustées, ces données nécessitent non seulement le développement de techniques d'exploration appropriées, mais aussi la mise en oeuvre d'abstractions pertinentes. De la même manière qu'un large éventail de cartes (géologiques, historiques, routières, topographiques, etc.) permettent de représenter différentes facettes de notre environnement géographique, un important panel d'abstractions est indispensable pour capturer les différents axes de complexité d'un système.

Le travail d'abstraction commence par le choix de la perspective que l'on souhaite privilégier. Autrement dit, quelles sont les qualités du système que l'on désire mettre en avant dans le modèle ? Ou, de manière inverse, quels sont les détails du système que l'on estime peu pertinents eu égard à l'angle d'attaque adopté ? Ces deux dernières décennies ont vu l'émergence de langages de modélisation de haut-niveau qui font bien sentir cette problématique. Qu'il s'agisse d'UML dans un contexte informatique, de SysML dans le secteur de la systémique ou de SBML pour la biologie, ces formalismes englobent différentes abstractions complémentaires. Le choix de l'une ou l'autre dépend de la finalité visée.

Ces abstractions sont primordiales à toutes les étapes d'étude d'un système, qu'il s'agisse de sa conception (pour un système que l'on cherche à définir), de son analyse (pour un système déjà existant et/ou que l'on souhaite améliorer) ou de son contrôle. Dans ce manuscrit, nous illustrerons l'enjeu majeur que représente la définition d'abstractions pertinentes, notamment dans le champ de la biologie des systèmes.

## **3.2 Distance entre système et modèle**

La construction même d'une représentation du réel induit de nombreux biais : méprise quant à l'interprétation d'un phénomène, limite de l'abstraction considérée, simple erreur d'étourderie, incompréhension d'une spécification exprimée en langage naturel, le processus de modélisation peut être entaché d'imprécisions dévastatrices. Dès la constitution d'un modèle, il importe donc de garder en tête qu'il diffère inévitablement du système réel.

On peut recourir à un modèle pour expliquer ou faciliter la compréhension d'un phénomène. Un modèle peut très bien répondre à cet objectif tout en étant intrinsèquement faux. En physique,

par exemple, le modèle atomique de Bohr est l'un des premiers enseignés aux lycéens alors qu'il peut être considéré comme faux et qu'un pan entier de recherche a vu le jour depuis pour traiter la mécanique atomique en profondeur grâce à la physique quantique. D'ailleurs, ces mêmes lycéens deviennent généralement sensibles aux limites de la modélisation en général quand ils sont confrontés, à travers moult expériences, à la dualité onde-corpuscule. Intrinsèquement, un modèle est une approximation du réel.

Le raffinement progressif d'un modèle pour en améliorer la fidélité par rapport au réel est une idée clef à laquelle nous nous raccrocherons dans ce manuscrit. Elle promet d'être cruciale dans le domaine émergent qu'est la biologie des systèmes. En effet, les modèles biologiques que nous construisons se basent sur les données accessibles aux biologistes à un instant  $t$ . Or, les techniques d'obtention de ces données se perfectionnant, de nouvelles informations viennent régulièrement enrichir les bases de données existantes, confirmant ou infirmant des connaissances précédemment établies. Face à de telles problématiques, nous proposons d'envisager deux approches différentes.

La première consiste à construire le modèle le plus général possible et à le raffiner progressivement avec les données biologiques jugées « sûres ». C'est la méthodologie que nous développerons dans la seconde partie de ce manuscrit : établir le modèle avec la plus large dynamique possible (appelée dynamique généralisée dans le cadre du formalisme de Frappes de Processus que nous avons introduit) et restreindre progressivement cette dynamique en y ajoutant des informations sur la coopération entre plusieurs composants et/ou la vitesse relative de différentes interactions. Une telle approche nécessite une certaine confiance dans les données dont nous disposons. C'est la raison pour laquelle nous nous orientons désormais vers une démarche complémentaire pour être en mesure de traiter des jeux de données éventuellement imparfaits et/ou amenés à être complétés au fil du temps. Cette approche est inspirée par les recherches dans le domaine de l'apprentissage automatique (ou *machine learning*). Nous en expliciterons l'intérêt dans le chapitre 5 de ce document, en présentant quelques travaux récents ainsi que des perspectives dans le domaine.

Indépendamment de la distance entre le réel et ses modèles, subsiste une problématique complémentaire, au cœur des mathématiques et de l'informatique : quel est le lien entre les différents formalismes de modélisation ? Depuis quelques années, des travaux salutaires ont été menés sur ce sujet, afin d'établir les liens entre différentes classes de modèles. Qu'il s'agisse de comparaisons (par exemple en termes de décidabilité et d'expressivité) ou de transformations de modèles, ce travail est primordial pour choisir le modèle le plus approprié, tant d'un point de vue théorique que pragmatique. Nous détaillerons cette problématique plus loin dans cette partie. Mais, avant, arrêtons-nous sur les spécificités de la modélisation en biologie.

### 3.3 Les enjeux de la modélisation en biologie

Les phénomènes de régulation constituent un facteur clé des systèmes biologiques, qu'on les étudie à l'échelle moléculaire ou écologique. Acquérir une compréhension précise des phénomènes de régulation est un des principaux objectifs de la biologie des systèmes. Cette discipline a émergé de la synergie entre la biologie cellulaire et la cybernétique [171] via la collaboration entre biologistes, physiciens et informaticiens [83]. L'approche de modélisation que nous avons choisie d'adopter se place dans la continuité de cet héritage : nous nous intéressons aux interactions entre les différents composants d'un système biologique et nous analysons comment ces interactions ont un impact sur le fonctionnement et le comportement du système dans son ensemble. Dans ce manuscrit, nous nous focalisons sur les applications dans l'univers de la génétique, mais le champ d'application de nos méthodes est bien plus large. Ainsi, le formalisme des Frappes de Processus que nous avons introduit peut s'appliquer aux systèmes concurrents (présentant plusieurs composants en interaction dont chacun ne peut, à un instant donné, n'être que dans un seul état), qu'il s'agisse d'un réseau de régulation biologique, d'une chaîne logistique ou d'un système embarqué.

Les progrès récents dans le domaine de la biologie moléculaire ont permis de dresser une carte complète du génome de nombreux organismes vivants. Dans le même temps, le développement des puces à ADN a donné accès aux séries temporelles d'expression de gènes de plusieurs milliers de gènes. Parallèlement, comme les interactions surviennent à différentes échelles (gènes, protéines, composants biochimiques, cellules, etc.), il est nécessaire de construire des méthodes qui peuvent formellement apprendre les données biologiques à un niveau systémique.

La modélisation des mécanismes de régulation biologique peut se décomposer en deux principales tendances. La première se fonde sur des équations différentielles ordinaires impliquant l'expression quantitative des composants en interaction. Malheureusement, ces équations sont généralement non-linéaires, ce qui rend impossible toute résolution analytique. De plus, les données résultant des expériences biologiques sont souvent bruitées, ce qui rend leur filtrage indispensable. De fait, même si la modélisation discrète peut être vue comme une abstraction moins fidèle, celle-ci s'est révélée efficace pour traiter beaucoup de questions biologiques, comme comprendre comment les systèmes biologiques évoluent ou déterminer si certains états sont accessibles. À terme, nous aimerions imaginer comment modifier l'évolution de ces systèmes afin d'éviter des comportements non-souhaités (pathologiques, par exemple).

### 3.4 Une large classe de modèles

Dans ce manuscrit, nous nous intéressons aux modèles discrets (et hybrides) permettant de modéliser des processus concurrents, notamment des phénomènes de régulation entre différents composants (biologiques) [52]. Mais avant de rentrer dans le détail de leurs différences au niveau sémantique, arrêtons-nous sur les principaux modèles dans le domaine.

### 3.4.1 Un aperçu des principaux modèles existants

#### 3.4.1.1 Réseaux booléens et discrets

Puisque, dans le domaine de la biologie, les données sont souvent plus qualitatives que quantitatives (au sens où le bruit des expériences est loin d'être négligeable), il est naturel qu'une vision alternative aux équations différentielles ait commencé à voir le jour à la fin des années 1960. Cette modélisation qualitative repose sur le principe suivant : l'expression d'un composant peut être encodée à l'aide d'une variable booléenne. Cela correspond au fait qu'un composant (par exemple un gène) est "allumé" (autrement dit : les régulations qu'il contrôle s'expriment) ou éteint. Les relations entre les variables booléennes sont gouvernées par des relations d'activation ou d'inhibition, respectivement représentées par des arcs signés positifs ou négatifs. Deux grandes familles de modèles ont développé ces préceptes : les réseaux booléens synchrones de Stuart Kauffman [90] d'une part, les réseaux de René Thomas asynchrones [168] d'autre part. Dans le paragraphe suivant, nous revenons sur les différences entre ces deux approches.

**Principe - Réseaux booléens et Réseaux de Thomas.** Nous revenons ici sur les formalismes de modélisation que nous avons évoqués dans le chapitre 2, à savoir les réseaux booléens conçus par Stuart Kauffman en 1969 d'une part, la description logique de René Thomas conçue à peu près à la même époque d'autre part. Même si la spécification détaillée de ces deux modèles est différente, ils reposent tous deux sur l'idée centrale est lié à l'état suivant par une fonction booléenne. La principale différence entre ces deux paradigmes réside en fait dans la fonction de mise à jour : alors que les réseaux booléens sont principalement étudiés avec une sémantique synchrone, l'approche de René Thomas considère une sémantique asynchrone. L'impact de cette hypothèse sur les propriétés dynamiques des systèmes a fait l'objet de plusieurs travaux récents [9, 130, 131].

Les recherches autour de ces modèles visent notamment à enrichir leur expressivité, identifier des sous-classes faciles à analyser, ou à développer de nouvelles méthodes efficaces d'études. Ainsi, dans le cas des réseaux booléens, Tatsuya Akutsu et al. ont dressé en 2008 un inventaire des algorithmes pour traiter l'inférence, l'identification des attracteurs et le contrôle [3]. Mais comme la taille de l'espace d'états croît de manière exponentielle avec le nombre de composants, les approches classiques sont rapidement inaptes à aborder de grands systèmes. C'est pourquoi un nombre important d'analyses ont été conçues pour rendre explicites les relations entre le graphe d'interaction du réseau et ses propriétés dynamiques. Par exemple, des études se sont concentrées sur les boucles de rétroactions, c'est-à-dire les chaînes circulaires d'interactions. Il est ainsi possible de montrer que le nombre d'interactions inhibitrices dans une boucle conduit directement à une condition nécessaire pour plusieurs états stationnaires [54]. Dans [143], les auteurs mettent en évidence ce lien entre le graphe d'interaction et les propriétés dynamiques des réseaux booléens. Dans [84, 87], les auteurs établissent une traduction directe des réseaux booléens en programmes logiques, et exploitent ainsi les méthodes permettant d'inférer des

programmes logiques pour calculer la trajectoire et les attracteurs des réseaux booléens.

En complément, un certain nombre de méthodes basées sur l'exploration de l'espace d'états ont été développées afin de vérifier un plus large éventail de propriétés : accessibilité, vivacité, identification des attracteurs, ... Toutefois, la plupart des approches classiques échouent à analyser des systèmes de grande taille. En effet, même en reposant sur des approches symboliques visant à accélérer l'analyse de l'espace d'états, celui-ci souffre de l'explosion combinatoire du nombre des états en fonction du nombre de composants du système.

En complément des travaux se focalisant sur la conception de représentations symboliques (par exemple inspirées des diagrammes de décision binaires [128]) de l'espace d'états, nous avons positionné notre contribution sur l'identification de sous-classes pertinentes de formalismes existants, sous-classes sur lesquelles il est possible d'établir des analyses beaucoup plus efficaces. C'est toute la spécificité du formalisme des Frappes de Processus, que nous présenterons à partir de la section suivante. Mais parce que nous n'avons pas souhaité nous arrêter à des formalismes uniquement booléens, présentons tout d'abord la raison de leurs extensions multi-valuées.

**Extensions multi-valuées.** Ces paradigmes booléens peuvent apparaître comme des modèles simplistes, mais ils ont conduit à des résultats significatifs sur le comportement des réseaux de régulation, notamment en matière de comportement cyclique ou d'états stationnaires. Mais les variables booléennes ne sont parfois pas suffisantes pour capturer la complexité intrinsèque de certains systèmes. Par exemple, quand un composant biologique est susceptible d'à la fois activer un gène et inhiber un troisième gène, il est très peu probable que ces interactions se déclenchent simultanément au même niveau de concentration du gène régulateur. L'intérêt de ces extensions multi-valuées a été discuté et illustré sur différents cas d'étude biologiques, tel le contrôle de l'immunité dans le phage-lambda [167] ou le réseau p53-Mdm2 [1]. De fait, les principes de modélisation qualitative ont été étendus de telle sorte que le modèle puisse capturer différents niveaux (discrets) d'expression. Cela a ainsi ouvert la voie aux réseaux discrets (ou logiques), étudiés à travers différents formalismes au cours des années, tels les réseaux de Petri [33]. Nous y reviendrons dans les prochaines sous-sections.

**Une autre extension : les multiplexes.** Sans améliorer l'expressivité théorique du modèle, mais en le rendant plus facile à manipuler, Bernot et al. ont introduit en 2008 la notion de multiplexes [20]. Ce formalisme distingue deux types de nœuds, d'une part ceux représentant les composants biologiques tels qu'habituellement modélisés comme nœuds dans les réseaux booléens ou discrets, d'autre part ceux qui modélisent une fonction booléenne (éventuellement multi-valuée) des nœuds en entrée. Cette fonction booléenne combine les régulateurs d'un composant et détermine ainsi si l'influence associée s'exerce positivement ou négativement sur le composant cible. Tout réseau représenté sous la forme de multiplexes peut être traduit en une "classique" représentation sous forme de réseau discret multi-valué ... au détriment toutefois d'une facilité de lecture.

### 3.4.1.2 Une dose de quantitatif

**Réseaux de Petri.** Ce panorama introductif ne serait pas complet sans citer notamment les travaux autour de l'application des réseaux de Petri (discrets et hybrides) dans le domaine de la biologie.

Les systèmes biologiques représentent l'essence même de systèmes dans lesquels la notion de concurrence est centrale. Ainsi, lors de la modélisation d'un réseau de régulation génétique, on doit représenter l'évolution respective de chacun des acteurs du système. Les réseaux de Petri constituent l'un des formalismes privilégiés pour représenter de manière concise cette notion de concurrence [145]. Différentes extensions de réseaux de Petri ont également été introduites pour modéliser et vérifier des systèmes temporisés ou hybrides. Dans le domaine de la biologie des systèmes, les réseaux de Petri ont été préférentiellement utilisés pour simuler le comportement des systèmes [34, 35]. Ce formalisme est associé à un certain nombre d'extensions, notamment les réseaux de Petri stochastiques qui peuvent permettre de représenter des comportements stochastiques, et constituent un outil de modélisation que les biologistes arrivent à appréhender facilement [22].

Les extensions stochastiques des réseaux de Petri s'avèrent efficaces, notamment, pour modéliser des systèmes biochimiques. Les principaux travaux portant sur les réseaux stochastiques concernent des modèles Markoviens [81]. Les modèles exponentiels ne sont malheureusement pas toujours suffisants. Certaines recherches ont donc visé les cas non-Markoviens. Dans [78], les auteurs s'intéressent à l'analyse de la réponse transitoire et des états stationnaires des réseaux de Petri stochastiques incorporant des activités concurrentes non-Markoviennes associées aux transitions. Pour ce faire, leur idée est de définir un modèle approché sur lequel ils peuvent mener une analyse exacte. Ils décrivent alors le comportement d'un modèle stochastique sous la forme de deux composants : une chaîne de Markov à temps continu et une chaîne de Markov en temps discret. Ainsi, dans ce paradigme, les activités non-Markoviennes ne sont prises en compte qu'à des instants discrets.

Des travaux menés en 2008 par Monika Heiner et al. ont montré comment les réseaux de Petri pouvaient être utilisés aussi bien pour des analyses qualitatives que quantitatives : dans [95], les auteurs proposent un formalisme qui dresse des liens entre les paradigmes qualitatifs, stochastiques et continus. Ils reconnaissent toutefois que des éléments de recherche plus approfondis sont nécessaires pour comprendre précisément les propriétés de chacune de ces descriptions. Plus récemment, depuis 2011, Claudine Chaouiya et al. ont investigué des formalismes de haut-niveau pour la représentation de systèmes biologiques via des réseaux de Petri [32]. Ces travaux se sont prolongés dans le domaine des analyses de toxicité [55].

**Extensions hybrides.** La dernière décennie a été le témoin d'un intérêt croissant pour la modélisation hybride des réseaux de régulation génétique. Par modélisation hybride, on entend l'intégration de délais quantitatifs dans la modélisation. Ces délais permettent de modéliser par

exemple le temps nécessaire entre le moment où une interaction devient effective et le moment où son effet se fait sentir, c'est-à-dire le moment où le gène régulé change de niveau qualitatif.

Les auteurs de [120] ont ainsi considéré, dans les années 2000, la modélisation des réseaux de régulation génétiques à travers les réseaux de Petri hybrides : l'avantage de cette extension hybride réside dans la possibilité de représenter des phénomènes biologiques, tels que le délai pour la transcription de l'ARN polymérase. Les avantages d'autres formalismes hybrides en biologie ont également fait l'objet de différentes études, qu'il s'agisse des automates temporisés [162] ou des automates hybrides [2].

Enfin, dans [44], les auteurs ont mis au point une extension directe de l'approche de modélisation discrète de René Thomas en introduisant des délais quantitatifs. Ces délais représentent le temps nécessaire pour un gène pour passer d'un niveau qualitatif discret au suivant (ou au précédent). Les avantages d'une telle modélisation ont été notamment illustrés sur un cas d'étude consistant en une analyse de la production de mucus dans la bactérie *Pseudomonas aeruginosa*.

### **3.4.2 Différence entre les modèles**

#### **3.4.2.1 Synchrones versus asynchrones**

L'une des premières questions lors de la construction d'un modèle consiste à savoir si deux événements peuvent avoir lieu simultanément (hypothèse synchrone) ou non (asynchrone). Ces deux familles d'approches se sont développées de manière concomitante car elles représentent des cas de concurrence assez différents. Par exemple, l'étude d'un système englobant des phénomènes se situant sur deux échelles temporelles, l'une très rapide et l'autre beaucoup plus lente, peut amener à considérer que certains phénomènes se déroulent simultanément et, de fait, à préférer l'hypothèse synchrone.

L'un des principaux intérêts des modèles synchrones est d'être généralement traitables d'un point de vue calculatoire, tandis que les algorithmes classiques d'exploration de l'espace d'états éprouvent des difficultés sur les modèles asynchrones. Pourtant, la modélisation synchrone repose des hypothèses assez contraignantes, à savoir que tous les composants biologiques peuvent opérer une transition simultanément et que cette transition nécessitera le même délai. Même si cela n'est habituellement pas réaliste d'un point de vue biologique, ces simplifications sont généralement suffisantes. Et ce, d'autant plus que la cinétique exacte des processus biologiques et l'ordre des transitions sont la plupart du temps inconnus. Cela dit, la sémantique asynchrone permet de rendre compte des comportements plus réalistes. À un instant donné, un unique gène peut ainsi voir son niveau d'expression évoluer. La conséquence, c'est que le modèle peut aboutir à une explosion combinatoire du nombre des états. Pour illustrer cela, les auteurs de [73] ont ainsi comparé le temps nécessaire pour calculer les attracteurs d'un large éventail de modèles (cellule de mammifère, , T-helper, cellule dendritique, ...) et ont comparé les résultats obtenus avec les sémantiques synchrones et asynchrones. Afin de rendre le calcul en sémantique asynchrone plus efficace, ils ont proposé un nouvel algorithme qui capture les attracteurs en rapprochant les

attracteurs des modèles asynchrones de leur équivalent synchrone. Ils ont identifié les cas où ces attracteurs étaient les mêmes, et ils ont implémenté leur approche dans un outil logiciel baptisé genYsis. Celui-ci aboutit à des performances meilleures que les autres approches existantes, parce qu'il tire justement profit d'une analyse combinée. Cet exemple montre tout l'intérêt d'une comparaison des résultats obtenus avec les modèles synchrones et ceux résultant des équivalents asynchrones. Non seulement peut-on établir des résultats théoriques intéressants, mais aussi peut-on améliorer les analyses existantes.

Il s'agit d'ailleurs d'un champ de recherches que d'autres, avant même les auteurs sus-cités, avaient entamé, notamment dans le cas des attracteurs. En effet, ces derniers donnent des informations précieuses sur les processus de différenciation cellulaire. Ainsi, plusieurs groupes ont montré que les attracteurs obtenus avec une sémantique synchrone [5, 122] et ceux qui résultent d'une sémantique asynchrone [21, 63] diffèrent.

Il faut donc prêter grand cas à cette problématique de sémantique, généralement résumée par le choix du type de mise à jour dans une approche fonctionnelle.

Pour nos travaux, nous avons choisi de considérer une sémantique asynchrone.

### **3.4.2.2 Déterminisme versus non-déterminisme**

De la même manière que la discussion autour du synchronisme et de l'asynchronisme, la construction d'un modèle conduit à s'interroger sur le déterminisme (ou non) associé aux phénomènes modélisés. Par non-déterminisme, nous entendons ici que, dans un état, un modèle peut être amené à évoluer selon des lignées différentes. Le non-déterminisme est généralement une hypothèse clef des systèmes à événements discrets. Il s'avère fréquent quand nous n'avons qu'une connaissance partielle d'un système, ce qui est souvent le cas avec l'étude de systèmes biologiques.

Considérons ainsi l'évolution d'un modèle qui, partant d'un état, aboutit à plusieurs états différents. Cet indéterminisme d'apparence peut trouver son origine dans plusieurs approximations :

- Une connaissance partielle du système : par exemple, les variables observées (voire les variables observables) ne rendent pas compte de manière suffisamment précise de l'état du système. C'est le cas de deux exécutions qui, en apparence, partent du même état mais qui, en fait, démarrent de deux états différents.
- Une volonté d'abstraction de la part du modélisateur : l'exercice de modélisation impose naturellement d'opérer des choix dans les dimensions caractéristiques d'un système. De fait, le comportement déterministe d'un système peut conduire à un indéterminisme dans le modèle qui lui est associé. Cette approximation à l'heure de la modélisation peut néanmoins être judicieuse, notamment dans une volonté de simplifier certaines fonctions du système.
- Enfin, le comportement même d'un système peut gouverner le choix de cet indéterminisme (au sens où il n'apparaît pas de principe de causalité dans son évolution) dans le modèle



associé. Il est des modèles (par exemple, la bascule en cycle lytique ou lysogénique du phage lambda) pour lesquels l'indéterminisme est une propriété clef. Quand on sait pertinemment que, partant d'un état donné, un système peut évoluer vers (au moins) deux états différents, l'indéterminisme du modèle permet de rendre compte avec fidélité ce type de comportement.

Parfois, bénéficiant d'informations supplémentaires sur un système, il est souhaitable de lever partiellement l'indéterminisme d'un modèle. Pour cela, il est possible de raffiner le modèle (par exemple, en ajoutant des arcs inhibiteurs pour régler les conflits dans un réseau de Petri).

### 3.4.2.3 Effet mémoire ou non (système non-markovien ou markovien)

Il existe un autre critère permettant d'ajuster la fidélité d'un modèle au système qu'il représente, à travers le choix d'une sémantique markovienne ou non-markovienne.

Les processus de Markov ont d'abord vu le jour dans le domaine des systèmes à événements discrets. L'hypothèse centrale, dans un processus de Markov, réside dans le fait que, partant d'un état, la prédiction du futur d'un processus ne dépend que de cet état. Autrement dit, la mémoire des événements passés n'apporte aucune information supplémentaire. Toute l'information utile pour la prédiction du futur du processus est contenue dans l'état courant du processus. De fait, on dira qu'un processus est markovien si son comportement à l'instant  $t + 1$  ne dépend que de son comportement à l'instant  $t$ . Évidemment, les processus de Markov ont été étendus au temps continu, en associant une loi de probabilités aux différents événements possibles.

Beaucoup de travaux, aussi bien en mathématiques, économie (étude des processus financiers) que biologie, font l'hypothèse que tous les processus en jeu sont markoviens. Ce modèle a connu beaucoup d'extensions ou de généralisations, à l'instar des processus semi-markoviens ou des processus de Markov cachés.

A contrario, il existe également des modèles nécessitant la prise en compte de l'historique des événements. Ainsi les modèles non-markoviens font l'objet d'abondantes études. Certains de ces modèles sont réductibles à des modèles markoviens ([80]). Les autres nécessitent des méthodes d'analyse qui leur sont propres. On retrouve des modèles non-markoviens dans de multiples domaines, tels la physique (le mouvement brownien), l'économie (les indices boursiers) ou encore la chimie architecturale (où le repliement d'une chaîne de polymères peut être représentée via une marche aléatoire auto-évitante, c'est-à-dire une marche aléatoire normale dans laquelle il est interdit de repasser deux fois au même endroit). La modélisation de mécaniques de régulation (par exemple génétique) n'échappe pas à ce type de questions. Récemment, nous avons vu émerger des travaux autour d'une modélisation des mécaniques de régulation génétique prenant en compte le passé. Nous avons nous aussi pris part à ce questionnement, comme en témoignera le chapitre 5, dans lequel l'état futur d'un réseau booléen ne dépend plus seulement de l'état présent, mais aussi des états passés (en remontant potentiellement l'historique du modèle de  $k \in \mathbb{N}$  pas,  $k$  restant évidemment inférieur au nombre de transitions opérées jusque là par le modèle).

Lors de la modélisation, le choix d'une sémantique markovienne ou non découle de deux aspects complémentaires : la connaissance disponible sur le système (est-on sûr d'avoir accès aux valeurs de toutes les variables signifiantes pour ce système?) et la complexité que l'on souhaite attacher au modèle résultant. Si on estime que l'état d'un modèle englobe suffisamment d'information pour permettre une prédiction parfaite du comportement du modèle depuis cet état, alors on optera pour une sémantique markovienne. A contrario, on pourra choisir un modèle non-markovien, en ajustant éventuellement la taille de la mémoire prise en considération, pour représenter le comportement d'un système pour lequel on sait manquer d'information.

#### **3.4.2.4 Discussion : quelle sémantique choisir ?**

Synchronisme, déterminisme, propriété de Markov, ... les caractéristiques dont nous avons discuté dans cette section sont intimement liées, et leur bonne compréhension est indispensable pour travailler de concert avec des biologistes à l'établissement de nouveaux modèles.

Donnons-en ici une illustration en mettant en évidence l'intérêt, par exemple, d'une sémantique non-markovienne. Ce type de modélisation peut en effet être utile, entre autres, pour évaluer si le comportement d'un système répond, ou non, à une forme de déterminisme. En effet, un système peut paraître indéterministe d'un point de vue atomique (c'est-à-dire que, partant d'un état donné, il admet deux successeurs directs qui sont différents), mais s'avère déterministe si on considère l'historique de son évolution. Ainsi, partant d'un modèle atomiquement indéterministe, l'étude de son comportement via un modèle non-markovien mettra en évidence un déterminisme à l'échelle de  $k$  pas temporels. Ce type d'information est utile pour obtenir une compréhension fine du système.

Étant données les différentes notions que nous avons abordées dans cette sous-section, nous avons choisi de concentrer notre travail sur des modèles asynchrones, généralement multi-valués, sans effet mémoire. Nous avons toutefois également entamé des travaux sur des sémantiques un peu différentes, que nous afficherons explicitement. En particulier, nous présenterons dans le chapitre 5 de ce manuscrit quelques recherches récentes visant à étendre les résultats existants dans le domaine des réseaux booléens à des modèles intégrant un effet mémoire (que nous appellerons Markov( $k$ )). Les résultats portant ainsi sur des modèles dotés de sémantiques un peu différentes seront ainsi explicitement mentionnés.

### **3.4.3 Comment choisir un modèle ?**

#### **3.4.3.1 Les problématiques d'analyse des modèles**

Grâce aux approches qualitatives, il est plus facile — même si cela reste assez compliqué — d'avoir une compréhension profonde des interactions en jeu dans un réseau de régulation. Cette méthodologie permet d'expliquer comment les composants d'un tel réseau interagissent ou sont éventuellement contrôlés par d'autres. Elle ouvre la voie à la prédiction de comportements jusque

là non-observés, et ainsi à la génération automatique de nouvelles expériences permettant de les confirmer (ou de les infirmer, auquel cas il est alors nécessaire de reconsidérer le modèle). Ceci est d’une importance cruciale, particulièrement dans le domaine de la biologie synthétique et de la conception de médicaments. Les défis rencontrés dans l’analyse et la prédiction de comportements tombent généralement dans l’une des quatre catégories suivantes :

- L’identification de paramètres : les modèles discrets tels que les réseaux booléens ou les réseaux de Thomas nécessitent non seulement de l’information sur la topologie du réseau (ce qui est habituellement considéré comme une connaissance préalable) et le type d’influences — activation ou inhibition — entre les composants, mais aussi la force respective de chacune des interactions. Cette donnée est nécessaire pour déterminer, par exemple, le comportement du modèle quand deux gènes ont des effets opposés sur une seule et même cible. Cette information permet de clarifier soit la fonction booléenne de mise à jour (réseau booléen), soit la paramétrisation du réseau de Thomas. Elle est obtenue à partir des observations biologiques expérimentales par model-checking [155] ou programmation par contraintes [71].
- L’inférence du modèle : plus global que le problème précédent, l’inférence du modèle ne se limite pas à des considérations sur les paramètres de celui-ci, mais embrasse également la structure du réseau. Étant donné des données de séries temporelles d’expression, traduites en système (généralement partiel) états-transitions, l’enjeu est alors de construire un modèle discret (voire temporisé) qui est cohérent avec les données disponibles.
- L’analyse de propriétés sur les modèles, telle que l’identification de ses états stationnaires, de ses attracteurs, de propriétés d’accessibilité, etc.
- Le contrôle du modèle : ce dernier point représente en quelque sorte le Graal de la biologie synthétique, notamment dans une perspective de thérapie génique. Il s’agit d’être en mesure d’enrichir le réseau existant avec un superviseur capable de contrôler le modèle de telle sorte qu’il satisfasse certaines propriétés souhaitées. Ou, a contrario, qu’il garantisse de ne jamais tomber dans un comportement non-souhaité.

### 3.4.3.2 Compromis entre expressivité et décidabilité

Généralement, l’*expressivité* d’un modèle (nous entendons par là sa capacité à représenter un nombre important de caractéristiques du système) et sa simplicité en termes de vérification (décidabilité et complexité algorithmique) s’opposent. Néanmoins, il est souvent important de choisir un modèle *suffisamment* expressif car nous évitons ainsi le pessimisme éventuel d’un modèle surévaluant le comportement réel du système. Ce pessimisme peut en effet avoir deux effets pervers : d’une part, conduire à considérer une propriété fausse sur le modèle alors qu’elle est vraie sur le système réel et, d’autre part, mener à un espace d’états infini alors que celui d’un modèle plus fin serait fini. Le modèle discret d’une application peut notamment aboutir à un

espace d'états infini alors qu'un modèle plus précis, enrichi des caractéristiques temporelles de l'application, bornera le nombre d'états du modèle (en restreignant l'ensemble des comportements possibles) et en facilitera la vérification.

De plus, il est nécessaire, pour vérifier certains types d'application (par exemple, celles faisant intervenir des problématiques d'ordonnancement de tâches), de prendre en compte la durée séparant certaines transitions ou actions. Pour ce faire, le temps *logique*, habituellement capturé par des modèles de type automate fini, n'est plus suffisant. En biologie, le rythme circadien est l'une des meilleures illustrations de ce besoin d'un modèle capturant le temps *chronométrique*. L'intégration de délais quantitatifs est en effet indispensable à l'analyse du cycle d'(environ) 24 heures caractérisant ce modèle [43].

#### 3.4.4 Limites des modèles existants

Dans nos travaux de thèse, nous nous étions focalisés sur des formalismes temporisés [104]. Ces approches, très expressives, sont particulièrement pertinentes pour étudier des systèmes tels que l'horloge circadienne. Seulement, elles peinent à aboutir à des résultats pratiques dès que le système est composé d'une dizaine de composants en interaction. Pour le comprendre intuitivement, il faut comprendre qu'un tel système, s'il est modélisé sous la forme d'un réseau de Petri temporel, implique plusieurs dizaines de transitions temporelles simultanément sensibilisées. Ce qui conduit inmanquablement aux limites des model-checkers traditionnels [108].

En outre, même si les données produites par les biologistes sont de plus en plus nombreuses et exhaustives, il subsiste un manque de précision dans l'obtention de données cinétiques quantitatives précises. Comme ces données expérimentales sont généralement bruitées, il est difficile d'inférer les valeurs des paramètres temporels associés. La plupart des approches développées jusqu'à maintenant dans une perspective d'apprentissage se limitent ainsi à une étude chronologique [163]. Il en résulte que l'identification des paramètres nécessite un raisonnement indirect, qui peut devenir délicat quand le modèle est constitué de plus de dix composants en interaction. De fait, il est dès lors extrêmement difficile de traiter des réseaux de régulation réalistes et de grande taille.

De fait, sans avoir abandonné totalement ce type de modélisation [8], nous avons souhaité - et c'est le cœur de la contribution présentée dans ce manuscrit - développer une approche différente, dont l'objectif est de traiter des systèmes de grande taille. Autrement dit, nous avons souhaité mettre en exergue une sous-classe de formalismes existants, sous-classe sur laquelle nous avons établi des analyses efficaces. C'est ce formalisme que nous présentons dans la prochaine section. Nous détaillerons les approches associées dans le prochain chapitre, en reliant celles-ci aux problématiques biologiques que nous avons souhaité résoudre.

### 3.4.5 Notre contribution : les Frappes de Processus (ou pourquoi un autre modèle)

#### 3.4.5.1 Principe

Le formalisme des Frappes de Processus a été introduit à travers la thèse de Loïc Paulevé, soutenue en 2011. Il vise notamment à proposer une réponse au problème de passage à l'échelle [139]. Cette approche qualitative repose sur trois concepts clefs (notamment inspirés par le  $\pi$ -calcul [23, 123]) :

- Les composants biologiques (e.g., par exemple les gènes) sont abstraits sous la forme de ce que nous avons appelé *sortes* (en effet, celles-ci sont plus générales que des gènes ou des composants "réels" au sens où elles pourront aussi représenter des artifices de modélisation permettant de capturer l'état de certaines fonctions booléennes ou multi-valuées au sein du système). Ces sortes sont, à leur tour, divisées en différents "états" (que l'on appellera "processus" par analogie avec le  $\pi$ -calcul) qui correspondent aux différents niveaux discrets qualitatifs des composants représentés par les sortes.
- Chaque interaction entre deux composants est modélisée de manière atomique par une *frappe* partant du processus d'une sorte vers un autre processus pour le faire bondir à un nouveau processus (appartenant à la même sorte que le processus frappé). Autrement dit, les frappes permettent de représenter la dynamique du système, au niveau le plus atomique possible (c'est-à-dire en représentant les interactions entre les composants deux à deux), sans interdire qu'un composant évolue lui-même sous son action propre (alors représentée via une auto-frappe, soit une frappe d'un processus sur lui-même pour le faire bondir à un autre processus).
- Une telle représentation permet de construire la dynamique généralisée la plus large possible, puis de procéder par raffinements successifs, par exemple à travers l'introduction de *sortes coopératives* qui représentent l'influence combinée de plusieurs gènes sur une seule et même cible.

Constituant une restriction particulière [141] des réseaux d'automates asynchrones et des réseaux de Petri saufs [19], les Frappes de Processus peuvent être utilisées pour étudier des systèmes dynamiques complexes avec un nombre important de composants en interactions, chacun de ces composants étant décrit par un faible nombre d'états internes.

Les Frappes de Processus se prêtent ainsi bien à la modélisation de réseaux de régulation génétique avec plusieurs niveaux d'abstraction. En considérant les interactions au niveau le plus atomique possible, elles ouvrent la voie à des méthodes d'analyse statique et d'interprétation abstraite pour étudier des propriétés dynamiques complexes. Nous illustrerons l'efficacité de cette approche, en montrant qu'elle peut être appliquée à des réseaux de taille importante (jusqu'à plusieurs centaines de composants). Nos méthodes produisent ainsi des réponses très rapidement pour la décision de propriétés d'accessibilité successives, là où des techniques bien

établies de model-checking symboliques échouent à cause de l'explosion de l'espace d'états. Grâce à cette efficacité, les Frappes de Processus ouvrent la voie à des perspectives prometteuses pour l'inférence de réseaux de régulation génétique lorsqu'un nombre important de données biologiques doivent être traitées. Nous détaillerons ces résultats ainsi que leurs perspectives dans les prochains chapitres.

Mais avant, nous allons revenir sur la définition des Frappes de Processus standards telles que nous les avons formalisées pour la première fois dans [139].

### 3.4.5.2 Présentation intuitive

En tant que restrictions du  $\pi$ -calcul [23, 123], les *Frappes de Processus standards* telles que données à la Définition 3.1, aussi appelées *Process Hitting*, ou plus simplement *Frappes de Processus* dans ce chapitre, permettent une modélisation atomique et asynchrone des interactions entre composants.

Un modèle de Frappes de Processus standards comporte un nombre fini de *sortes* généralement notées  $a, b, c, \dots$ . Celles-ci permettent de représenter les différentes entités du modèle, qu'il s'agisse de composants ayant une réalité biologique (gène, protéine...) ou d'entités nécessaires à la modélisation (comme les sortes coopératives qui seront décrites à la section 3.4.5.5). Chaque sorte contient plusieurs *processus*, qui représentent les différents niveaux d'expression discrets accessibles par la sorte, et qui sont notés  $a_i$  où  $a$  est le nom de la sorte et  $i$  l'indice du processus dans cette sorte. Un processus n'appartient qu'à une unique sorte. Chaque processus est dit *actif* (on écrira aussi : *présent*) s'il représente le niveau d'expression dans lequel doit se trouver sa sorte à un certain moment. Un *état* du modèle est donc décrit par l'ensemble des processus actifs à un instant donné, avec exactement un processus actif par sorte — afin de ne pas sur-représenter ou sous-représenter le niveau d'expression courant d'une sorte.

La dynamique est introduite dans les Frappes de Processus par des *actions* qui permettent de modifier le processus actif d'une sorte, à la condition éventuelle qu'un processus donné d'une autre sorte soit présent. Une action consiste donc en un triplet de processus  $a_i \rightarrow b_j \uparrow b_k$  qui se lit : «  $a_i$  frappe  $b_j$  pour le faire bondir en  $b_k$  », et qui signifie que si, dans un état donné, les processus  $a_i$  et  $b_j$  sont tous les deux actifs, alors il est possible d'activer  $b_k$  (et de désactiver  $b_j$ ) dans l'état suivant. Autrement dit, le processus actif de la sorte  $b$  peut *bondir* de  $b_j$  à  $b_k$  à condition que  $a_i$  soit actif ; lorsque cela arrive, on dit qu'on a *joué* l'action  $a_i \rightarrow b_j \uparrow b_k$ . Par convention, on contraint de plus que  $b_j \neq b_k$  pour assurer que le jeu d'une action provoque bien le changement d'un processus actif. Il est aussi possible de définir une auto-action, où  $a_i = b_j$  (et nécessairement  $a = b$ ), qui permet de représenter le cas particulier où le processus  $b_j$  peut bondir en  $b_k$  sans autre condition.

Les Frappes de Processus sont conçues comme un formalisme à temps discret asynchrone, ce qui signifie que l'évolution dans un tel modèle est modélisée par une succession de pas de temps discrets qui représentent la succession des états du modèle, et exactement une action est

jouée entre deux états successifs. Cela implique qu'un seul processus actif à la fois peut bondir entre deux pas de temps successifs, et donc qu'une seule sorte peut évoluer entre deux états. De plus, cela rend la dynamique des Frappes de Processus indéterministe, car à tout état du modèle peuvent correspondre plusieurs états successeurs dans le cas où plusieurs actions peuvent y être jouées. Enfin, nous notons que si aucune action n'est jouable dans un état, alors celui-ci ne possède pas de successeur et le modèle ne peut plus évoluer.

**Définition 3.1** (Frappes de Processus standards). Les *Frappes de Processus standards* sont définies par un triplet  $\mathcal{PH} = (\Sigma; \mathcal{L}; \mathcal{H})$ , où :

- $\Sigma \stackrel{\text{def}}{=} \{a, b, \dots\}$  est l'ensemble fini et dénombrable des *sortes* ;
- $\mathcal{L} \stackrel{\text{def}}{=} \bigotimes_{a \in \Sigma} \mathcal{L}_a$  est l'ensemble fini des *états*, où  $\mathcal{L}_a = \{a_0, \dots, a_{l_a}\}$  est l'ensemble fini et dénombrable des *processus* de la sorte  $a \in \Sigma$  et  $l_a \in \mathbb{N}^*$ , chaque processus appartenant à une unique sorte :  $\forall (a_i; b_j) \in \mathcal{L}_a \times \mathcal{L}_b, a \neq b \Rightarrow a_i \neq b_j$  ;
- $\mathcal{H} \stackrel{\text{def}}{=} \{a_i \rightarrow b_j \uparrow b_k \mid (a; b) \in \Sigma \times \Sigma \wedge (a_i; b_j; b_k) \in \mathcal{L}_a \times \mathcal{L}_b \times \mathcal{L}_b \wedge b_j \neq b_k \wedge a = b \Rightarrow a_i = b_j\}$  est l'ensemble fini des *actions*.

On note  $\mathbf{Proc} \stackrel{\text{def}}{=} \bigcup_{a \in \Sigma} \mathcal{L}_a$  l'ensemble de tous les processus. La sorte d'un processus  $a_i$  est donnée par  $\text{sorte}(a_i) = a$  ; on définit aussi l'ensemble des sortes d'une action ou d'un ensemble de processus par :

$$\forall h \in \mathcal{H}, \text{sortes}(h) = \{\text{sorte}(\text{frappeur}(h)), \text{sorte}(\text{cible}(h))\}$$

$$\forall A \subset \mathbf{Proc}, \text{sortes}(A) = \{\text{sorte}(p) \mid p \in A\}$$

Étant donné un état  $s \in \mathcal{L}$ , le processus de la sorte  $a \in \Sigma$  présent dans  $s$  est donné par  $s[a]$ , c'est-à-dire la coordonnée correspondant à  $a$  dans l'état  $s$ . Si  $a_i \in \mathcal{L}_a$ , nous définissons la notation :  $a_i \in s \stackrel{\text{def}}{\Leftrightarrow} s[a] = a_i$  ; par extension, si  $ps \subset \mathbf{Proc}$ , on écrit alors :  $ps \subseteq s \stackrel{\text{def}}{\Leftrightarrow} \forall p \in ps, p \in s$ . Pour toute action  $h = a_i \rightarrow b_j \uparrow b_k \in \mathcal{H}$ ,  $a_i$  est appelé le *frappeur*,  $b_j$  la *cible* et  $b_k$  le *bond* de  $h$ , et on note :  $\text{frappeur}(h) = a_i$ ,  $\text{cible}(h) = b_j$  et  $\text{bond}(h) = b_k$ .

### 3.4.5.3 Exemple jouet

L'exemple adopté ici est repris du manuscrit de la thèse de Maxime Folschette, soutenue en octobre 2014 [64].

*Exemple.* La figure 3.1 page 50 illustre une représentation possible des Frappes de Processus standards. Le modèle  $\mathcal{PH} = (\Sigma, \mathcal{L}, \mathcal{H})$  représenté comporte trois sortes :  $\Sigma = \{a, c, f\}$ . Chaque sorte comporte exactement deux processus :

$$\mathcal{L}_a = \{a_0, a_1\} ; \quad \mathcal{L}_b = \{b_0, b_1\} ; \quad \mathcal{L}_f = \{f_0, f_1\} .$$

On peut notamment en déduire le nombre total d'états du système :  $|\mathcal{L}| = |\mathcal{L}_a| \cdot |\mathcal{L}_c| \cdot |\mathcal{L}_f| = 2^3 = 8$ . Cette grandeur n'est cependant donnée qu'à titre indicatif, car nous évitons de construire

explicitement l'espace des états dont la taille est exponentielle dans le nombre de sortes et de processus du modèle. Enfin, le modèle étudié comporte 7 actions :

$$\mathcal{H} = \left\{ \begin{array}{ll} c_1 \rightarrow a_1 \uparrow a_0 & , \quad c_0 \rightarrow a_0 \uparrow a_1 & , \\ c_1 \rightarrow c_1 \uparrow c_0 & , \quad f_1 \rightarrow c_0 \uparrow c_1 & , \\ f_1 \rightarrow a_0 \uparrow a_1 & , \quad f_0 \rightarrow c_1 \uparrow c_0 & , \\ f_1 \rightarrow f_1 \uparrow f_0 & \end{array} \right\}$$

Dans cet exemple, les Frappes de Processus représentent un modèle simplifié du mécanisme de segmentation des métazoaires qui permet par exemple de décrire la production de rayures chez les drosophiles. Il a été originellement établi *in silico* en 2007 par [70] à l'aide d'un formalisme à base d'équations différentielles, et le modèle présenté ici est inspiré du modèle proposé par [139].

Les trois sortes  $a$ ,  $b$  et  $c$  de ce modèle représentent différents gènes du système, que nous qualifierons d'*actifs* dans le reste de ce document lorsqu'ils seront à l'état 1 (ce qui ne doit pas être confondu avec le caractère *actif* d'un processus). La production de pigment est déclenchée par le produit du gène  $a$ , et une succession d'activations et de désactivations de celui-ci permet donc de produire des rayures. Pour que celles-ci soient régulières, il est donc nécessaire que la durée d'activation de  $a$  soit constante, et que la durée entre deux activations le soit aussi. Ce mécanisme est réglé par le gène  $c$  qui inhibe à la fois le gène  $a$  à intervalles réguliers, et s'inhibe lui-même afin d'avoir le rôle d'une horloge. Enfin, le procédé complet est dirigé par le gène  $f$  qui, lorsqu'il est actif, permet la progression d'un front au niveau duquel les pigments sont déposés. Ce gène peut aussi s'auto-inhiber après une certaine période, faisant cesser l'oscillation de l'horloge et ainsi la production de rayures.

#### 3.4.5.4 Notions complémentaires

Les séquences d'actions ont un rôle particulier pour les Frappes de Processus. Elles permettent notamment d'abstraire une dynamique locale en se concentrant sur la conséquence plutôt que sur la cause, et seront notamment utiles pour la méthode d'analyse statique développée au chapitre 4. Pour toute séquence d'actions  $A$ , on note  $\text{sortes}(A)$  l'ensemble des sortes dont au moins un processus figure dans  $A$  en tant que frappeur, cible ou bond d'une action. De plus, pour toute sorte  $a$ , on note  $\text{premier}_a(A)$  le premier processus de  $a$  référencé dans  $A$ , en tant que frappeur ou en tant que cible, et  $\text{dernier}_a(A)$  le dernier, en tant que frappeur ou en tant que bond. On note en conséquence  $\text{support}(A)$  l'ensemble de tous ces premiers processus, et  $\text{fin}(A)$  l'ensemble de tous ces derniers processus.



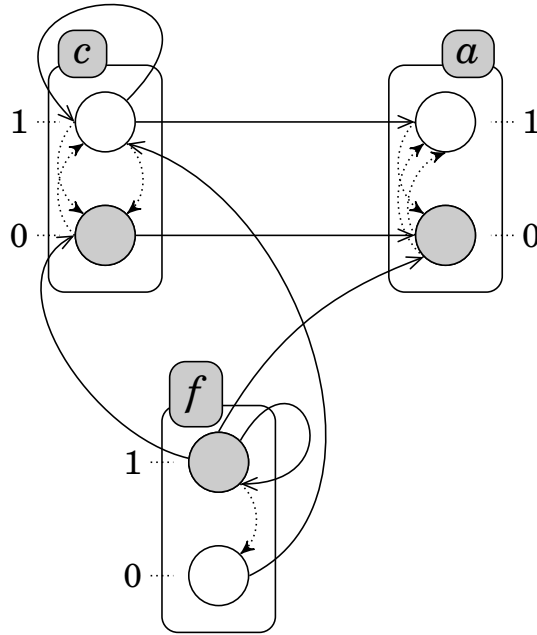


FIGURE 3.1 – Un exemple de Frappes de Processus standards. Les sortes sont représentées par des rectangles arrondis contenant des cercles représentant les processus. Ainsi, le processus  $a_1$  est représenté par le cercle marqué « 1 » dans le rectangle étiqueté «  $a$  », etc. Chaque action est de plus symbolisée par un couple de flèches, l'une en trait plein et l'autre en pointillés ; par exemple, l'action  $c_1 \rightarrow a_1 \uparrow a_0$  est représentée par une flèche pleine entre les processus  $c_1$  et  $a_1$  suivie d'une flèche en pointillés entre  $a_1$  et  $a_0$ . Enfin, les processus grisés représentent un état possible pour ces Frappes de Processus :  $\langle a_0, c_0, f_1 \rangle$ , qui peut aussi faire office d'état initial pour ce modèle.

**Définition 3.2** (premier, dernier, support et fin).

$$\text{premier}_a(A) = \begin{cases} \emptyset & \text{si } a \notin \text{sortes}(A) \\ \text{frappeur}(A_m) & \text{si } m = \min\{n \in \mathbb{I}^A \mid a \in \text{sortes}(A_n)\} \\ & \wedge \text{sorte}(\text{frappeur}(A_m)) = a \\ \text{cible}(A_m) & \text{sinon si } m = \min\{n \in \mathbb{I}^A \mid a \in \text{sortes}(A_n)\} \\ & \wedge \text{sorte}(\text{cible}(A_m)) = a \end{cases}$$

$$\text{dernier}_a(A) = \begin{cases} \emptyset & \text{si } a \notin \text{sortes}(A) \\ \text{bond}(A_m) & \text{si } m = \max\{n \in \mathbb{I}^A \mid a \in \text{sortes}(A_n)\} \\ & \wedge \text{sorte}(\text{bond}(A_m)) = a \\ \text{frappeur}(A_m) & \text{sinon si } m = \max\{n \in \mathbb{I}^A \mid a \in \text{sortes}(A_n)\} \\ & \wedge \text{sorte}(\text{frappeur}(A_m)) = a \end{cases}$$

$$\text{support}(A) = \{p \in \mathbf{Proc} \mid \text{sorte}(p) \in \text{sortes}(A) \wedge p = \text{premier}_{\text{sorte}(p)}(\delta)\}$$

$$\text{fin}(A) = \{p \in \mathbf{Proc} \mid \text{sorte}(p) \in \text{sortes}(A) \wedge p = \text{dernier}_{\text{sorte}(p)}(\delta)\}$$

La Definition 3.3 établit la notion de sous-état sur un ensemble de sortes, c'est-à-dire un ensemble de processus qui sont deux à deux de sortes différentes, ce qui permet de ne considérer qu'une partie d'un état complet. Nous notons  $\mathcal{L}^\diamond$  l'ensemble de tous les sous-états et nous constatons qu'un état est *a fortiori* un sous-état :  $\mathcal{L} \subset \mathcal{L}^\diamond$ . Nous notons de plus  $\mathbf{Proc}^\diamond$  l'ensemble des sous-états désordonnés, c'est-à-dire dont l'ordre entre les sortes a été oublié. Le recouvrement d'un état  $s$  par un processus  $a_i$  est formalisé à la Definition 3.4 par un état identique à  $s$ , sauf pour le processus de  $a$  qui a été remplacé par  $a_i$ , ce qui permettra de définir la dynamique des Frappes de Processus à la Definition 3.7. La définition de recouvrement est aussi étendue à un sous-état désordonné, autrement dit, à un ensemble de processus contenant au plus un processus par sorte.

**Définition 3.3** (Sous-états ( $\mathcal{L}^\diamond$ )). Si  $S \subset \Sigma$  est un ensemble de sortes, un sous-état sur  $S$  est un élément de :  $\mathcal{L}_S^\diamond \stackrel{\text{def}}{=} \bigotimes_{a \in S} \mathcal{L}_a$ . L'ensemble de tous les sous-états est noté :  $\mathcal{L}^\diamond \stackrel{\text{def}}{=} \bigcup_{S \in \wp(\Sigma)} \mathcal{L}_S^\diamond$ . De plus, si  $\sigma \in \mathcal{L}^\diamond$  et  $s \in \mathcal{L}$ , on note alors :

$$\sigma \subseteq s \stackrel{\text{def}}{\Leftrightarrow} \forall a_i \in \mathbf{Proc}, a_i \in \sigma \Rightarrow a_i \in s .$$

Enfin, si  $S \subset \Sigma$ , on note :  $\mathbf{Proc}_S^\diamond = \{\widetilde{ps} \subset \mathbf{Proc} \mid ps \in \mathcal{L}_S^\diamond\}$  et  $\mathbf{Proc}^\diamond = \{\widetilde{ps} \subset \mathbf{Proc} \mid ps \in \mathcal{L}^\diamond\}$ .

**Définition 3.4** (Recouvrement ( $\mathbb{m} : \mathcal{L} \times \mathbf{Proc} \rightarrow \mathcal{L}$ )). Étant donné un état  $s \in \mathcal{L}$  et un processus  $a_i \in \mathbf{Proc}$ ,  $(s \mathbb{m} a_i)$  est l'état défini par :  $(s \mathbb{m} a_i)[a] = a_i \wedge \forall b \neq a, (s \mathbb{m} a_i)[b] = s[b]$ . On étend de plus cette définition à un ensemble de processus par le recouvrement de l'état par chaque processus, à condition que les processus de l'ensemble soient tous de sortes différentes :  $\forall ps \in \mathbf{Proc}^\diamond, s \mathbb{m} ps = s \mathbb{m}_{a_i \in ps} a_i$ .

Nous définissons dans la suite les outils nécessaires à la dynamique des Frappes de Processus. Pour cela, nous introduisons la notion de propriété de *jouabilité* à la Definition 3.5, qui est semblable à une formule booléenne dont les atomes sont des processus dans  $\mathbf{Proc}$ . Le langage des propriétés de jouabilité permet de décrire la présence d'une configuration de processus actifs dans un état donné, ce qui permet donc notamment de décrire en termes formels la jouabilité d'une action. Cela est mis en pratique pour les Frappes de Processus standards à la Definition 3.6 où nous définissons l'*opérateur de jouabilité* de ce formalisme, qui est une formule associant à chaque action sa propriété de jouabilité propre. Pour finir, la dynamique des Frappes de Processus est donnée à la Definition 3.7 à partir de celle d'opérateur de jouabilité. Les définitions de propriété de jouabilité (Definition 3.5) et de dynamique des Frappes de Processus (Definition 3.7) sont volontairement assez générales pour pouvoir être réutilisées ou pour définir la sémantique d'autres formalismes de Frappes de Processus.

**Définition 3.5** (Propriété de jouabilité ( $F$ )). Une *propriété de jouabilité* est un élément du langage  $F$  défini inductivement par :

- $\top$  et  $\perp$  appartiennent à  $F$  ;
- si  $a \in \Sigma$  et  $a_i \in \mathcal{L}_a$ , alors  $a_i$  appartient à  $F$  et est appelé un *atome* ;
- si  $P \in F$  et  $Q \in F$ , alors  $\neg P$ ,  $P \wedge Q$  et  $P \vee Q$  appartiennent à  $F$ .

Si  $P \in F$  est une propriété de jouabilité et  $\sigma \in \mathcal{L}^\diamond$  est un sous-état, on note  $[P](\sigma)$  l'évaluation de  $P$  dans  $\sigma$  :

- si  $P = a_i \in \mathcal{L}_a$  est un atome, avec  $a \in \Sigma$ , alors  $[a_i](\sigma)$  est vraie si et seulement si  $a_i \in \sigma$  ;
- si  $P$  n'est pas un atome, alors  $[P](\sigma)$  est vraie si et seulement si on peut l'évaluer récursivement comme vraie en utilisant la sémantique habituelle des opérateurs  $\neg$ ,  $\wedge$  et  $\vee$  et des constantes  $\top$  et  $\perp$ .

Une fonction  $F : \mathcal{H} \rightarrow F$  associant à toute action une propriété de jouabilité est appelée un *opérateur de jouabilité*.

Étant donné que ce langage n'utilise que des opérateurs logiques classiques, les propriétés de la logique booléenne sont applicables aux propriétés de jouabilité, à savoir celles concernant la distributivité, l'associativité et la commutativité, ainsi que les lois de De Morgan concernant la négation.

Il en résulte notamment la propriété suivante, permettant d'évaluer la négation d'un atome, et qui dérive naturellement du fait que si un processus n'est pas actif dans un état donné, cela signifie alors qu'un autre processus de la même sorte l'est :

$$\forall a \in \Sigma, \forall a_i \in \mathcal{L}_a, \forall \sigma \in \mathcal{L}^\diamond, [\neg a_i](\sigma) \Leftrightarrow \left[ \bigvee_{\substack{a_j \in \mathcal{L}_a \\ a_j \neq a_i}} a_j \right](\sigma)$$

L'opérateur de jouabilité  $F$  donné à la Définition 3.6 est propre aux Frappes de Processus standards. En revanche, la dynamique donnée à la Définition 3.7 est générale à toutes les Frappes de Processus, et peut donc à fortiori être utilisée avec l'opérateur  $F$  des Frappes de Processus standards.

**Définition 3.6** (Opérateur de jouabilité ( $F : \mathcal{H} \rightarrow F$ )). L'opérateur de jouabilité des Frappes de Processus est défini par :

$$\forall h \in \mathcal{H}, F(h) \equiv \text{frappeur}(h) \wedge \text{cible}(h) .$$

**Définition 3.7** (Dynamique des Frappes de Processus ( $\rightarrow_{\mathcal{PH}}$ )). Une action  $h \in \mathcal{H}$  est dite *jouable* dans l'état  $s \in \mathcal{L}$  si et seulement si :  $[F(h)](s)$ . Dans ce cas,  $(s \cdot h)$  est l'état résultant du jeu de l'action  $h$  dans  $s$ , et on le définit par :  $(s \cdot h) = s \mathbin{\mathbb{M}} \text{bond}(h)$ . De plus, on note alors :  $s \rightarrow_{\mathcal{PH}} (s \cdot h)$ .

Si  $s \in \mathcal{L}$ , un *scénario*  $\delta$  dans  $s$  est une séquence d'actions de  $\mathcal{H}$  qui peuvent être jouées successivement depuis  $s$ . L'ensemble de tous les scénarios dans  $s$  est noté **Sce**( $s$ ).

*Remarque.* Les Frappes de Processus standards possèdent une dynamique totalement asynchrone : entre deux états, une unique action peut être jouée. Ce choix de conception est largement inspiré du modèle de Thomas, dont la dynamique est aussi totalement asynchrone. Biologiquement, une telle hypothèse est cohérente avec le fait que la probabilité que deux composants franchissent simultanément un seuil d'expression est infiniment faible. Cependant, cette sémantique asynchrone permet aussi de simplifier la dynamique des Frappes de Processus standards en assurant que deux états successifs ne varient que d'un seul processus (en fait : exactement d'un processus). Cela a notamment permis le développement de l'analyse statique évoquée plus loin, au chapitre 4.

*Exemple.* La dynamique des Frappes de Processus standards de la figure 3.1 est représentée à la figure 3.2. Nous pouvons notamment y observer le comportement stationnaire normal du modèle qui consiste en une oscillation alternée des processus actifs des sortes  $a$  et  $c$  :

$$\langle a_0, c_0, f_1 \rangle \rightarrow_{\mathcal{PH}} \langle a_1, c_0, f_1 \rangle \rightarrow_{\mathcal{PH}} \langle a_1, c_1, f_1 \rangle \rightarrow_{\mathcal{PH}} \langle a_0, c_1, f_1 \rangle \rightarrow_{\mathcal{PH}} \langle a_0, c_0, f_1 \rangle \rightarrow_{\mathcal{PH}} \dots$$

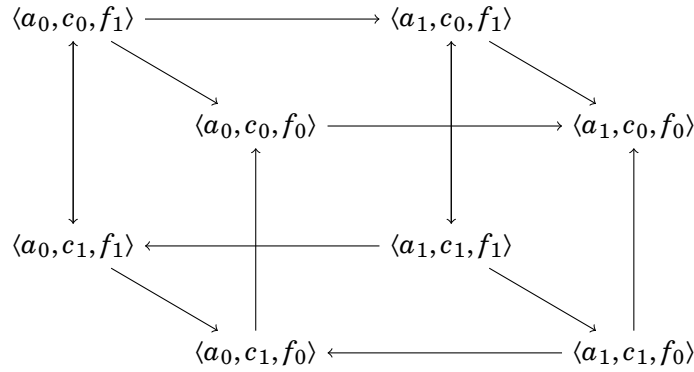


FIGURE 3.2 – Représentation de la dynamique du modèle de Frappes de Processus standards de la figure 3.1. Chaque état est représenté par un triplet  $\langle a_i, c_j, f_k \rangle$  où  $i, j$  et  $k$  représentent respectivement le niveau d'expression de  $a, c$  et  $f$ . Une transition entre deux états est représentée par une flèche.

### 3.4.5.5 Modéliser par raffinements successifs

Le principe de modélisation par les Frappes de Processus repose sur les principes suivants :

- La topologie du modèle (correspondant au Graphe des Interactions d'un réseau de Thomas par exemple) est supposée connue.
- Initialement, nous n'imposons aucune contrainte sur les fonctions logiques caractérisant les interactions entre les composants.

Il est dès lors possible de construire ce que nous appelons la *dynamique généralisée* du modèle. C'est-à-dire la dynamique la plus large possible, le but étant ensuite de restreindre cette dynamique en fonction des connaissances supplémentaires sur le système et/ou des hypothèses admissibles sur celui-ci.

Typiquement, les hypothèses faites à ce niveau sont, à l'instar de ce que nous avons proposé comme exemple dans [135], l'hypothèse de valeurs extrêmes, c'est-à-dire : pour chaque composant, si tous ses activateurs sont présents (resp. absents) et si tous ses inhibiteurs sont absents (resp. présents), le niveau du composant ne peut pas décroître (resp. augmenter).

**Coopérations.** Dans les Frappes de Processus, le changement d'état local d'une sorte est contrôlé par un unique processus, le *frappeur*. Dans ces conditions, comment encoder les changements locaux contrôlés par la présence d'au moins deux processus actifs. Par exemple, comment représenter le fait qu'un processus  $c_0$  ne bondit en  $c_1$  que si deux processus de deux sortes différentes,  $a_0$  et  $b_1$  sont actifs. Un tel comportement est appelée une *coopération* entre les sortes  $a$  et  $b$  pour agir sur  $c_0$ . Les coopérations sont typiquement spécifiées en utilisant des fonctions logiques [20, 155].

Afin de représenter la coopération entre composants - et, de fait, procéder au raffinement de la dynamique du modèle - nous avons proposé d'ajouter des sortes particulières appelées *sortes coopératives*, qui servent exclusivement à la modélisation [139]. Une sorte coopérative permet de représenter l'état conjoint de plusieurs sortes dans le modèle. Pour cela, à chaque processus de la sorte coopérative correspond un sous-état des sortes qu'elle *représente*. Ainsi, il est possible de représenter les différents états combinés d'un ensemble de sortes, afin de ne jouer une action que dans une configuration particulière. Ces sortes ont l'avantage d'être des sortes standards, et donc de ne pas nécessiter d'enrichissement particulier de la sémantique. De plus, leur utilisation n'a pas d'impact sur les méthodes d'analyse de la dynamique développées, et que nous présenterons dans le chapitre suivant. En effet, les performances de ces méthodes sont principalement dépendantes du nombre de processus dans chaque sorte, et non le nombre total de sortes.

**Exemple.** Les Frappes de Processus standards de la figure 3.3 reprennent les trois sortes  $f$ ,  $c$  et  $a$  du modèle de la figure 3.1 et comprennent en plus une sorte coopérative  $fc$  permettant de détecter la présence simultanée de  $f_1$  et  $c_1$ . Les processus de  $fc$  décrivent les combinaisons possibles des états de  $f$  et de  $c$  :  $fc_{00}$  correspond à  $f_0$  et  $c_0$ ,  $fc_{01}$  correspond à  $f_0$  et  $c_1$ , etc. Les actions en amont de cette sorte coopérative permettent de la mettre à jour, c'est-à-dire de changer son processus actif en fonction des évolutions du processus actif de  $f$  et  $c$ . Par exemple, si  $f_1$  est

actif, les deux actions  $f_1 \rightarrow fc_{00} \dot{\rightarrow} fc_{10}$  et  $f_1 \rightarrow fc_{01} \dot{\rightarrow} fc_{11}$  effectuent cette mise à jour en faisant bondir le processus actif de  $fc$  depuis un processus représentant la présence de  $f_0$  ( $fc_{00}$  ou  $fc_{01}$ ) vers le processus correspondant représentant la présence de  $f_1$  (respectivement  $fc_{10}$  ou  $fc_{11}$ ). Son action en aval,  $fc_{11} \rightarrow c_0 \dot{\rightarrow} c_1$ , joue alors le rôle d'une coopération entre  $f_1$  et  $c_0$  pour frapper  $a_0$  et le faire bondir en  $a_1$ .

*Exemple.* Afin de pallier partiellement l'absence de coopération entre les sortes  $f$  et  $c$  dans le modèle de la figure 3.1, il est possible d'intégrer la sorte coopérative décrite à la figure 3.3. La figure 3.4 propose un modèle corrigé de cette manière, avec une sorte coopérative  $fc$  permettant de détecter la présence de  $f_1$  et  $c_0$ . Les deux actions  $c_0 \rightarrow a_0 \dot{\rightarrow} a_1$  et  $c_0 \rightarrow a_0 \dot{\rightarrow} a_1$  sont alors remplacées par une action  $fc_{10} \rightarrow a_0 \dot{\rightarrow} a_1$  afin d'avoir une véritable coopération entre ces deux processus pour activer  $a$ .

Les sortes coopératives possèdent cependant une caractéristique qui peut s'avérer un inconvénient : le formalisme des Frappes de Processus standards étant totalement asynchrone et indéterministe, les actions sont totalement indépendantes. De fait, les sortes coopératives ne sont pas nécessairement mises à jour immédiatement après un bond du processus actif d'une des sortes qu'elles représentent. Il peut ainsi exister un « décalage temporel » entre le changement de processus actif d'une sorte et la mise à jour des sortes coopératives. Ce décalage temporel permet alors de jouer des actions modélisant des coopérations dans des états où la coopération n'est plus possible, car même si l'un des processus modélisant la coopération a bondi, la sorte coopérative peut ne pas en avoir fait de même. Or ce décalage peut aussi aboutir à des comportements indésirables. En effet, le processus actif d'une sorte coopérative ne correspond de fait pas à l'état courant des sortes représentées, mais uniquement à une combinaison d'états passés. Il est alors possible d'activer un processus de la sorte coopérative correspondant à un sous-état artificiel, c'est-à-dire non accessible aux sortes représentées.

La complexité de ces encodages est exponentielle en le nombre de sortes coopératives.

Cela dit, ces problèmes liés aux sortes coopératives nécessitent d'être relativisés à deux titres. D'une part, le principe de sortes coopératives mises à jour de manière différée peut s'avérer utile pour représenter des mécanismes de complexation. D'autre part, nous avons envisagé des extensions pour une mise à jour simultanée d'un ensemble de sortes, extensions que nous présentons au paragraphe suivant.

**Ajout de priorités.** Comme nous l'avons mis en évidence dans la sous-section précédente, la modélisation des coopérations avec des sortes coopératives présente l'inconvénient de sur-approximer la dynamique réelle du réseau. En effet, toutes les actions étant considérées sur un même plan, la sorte coopérative associée à une combinaison de sortes n'est pas forcément immédiatement mise à jour après que l'une de ces sortes a été mise à jour. Du fait de ce décalage temporel (chronologique) entre la mise à jour d'une sorte et la mise à jour de sa sorte coopérative associée, les analyses dynamiques associées sur-approximent la dynamique réelle.

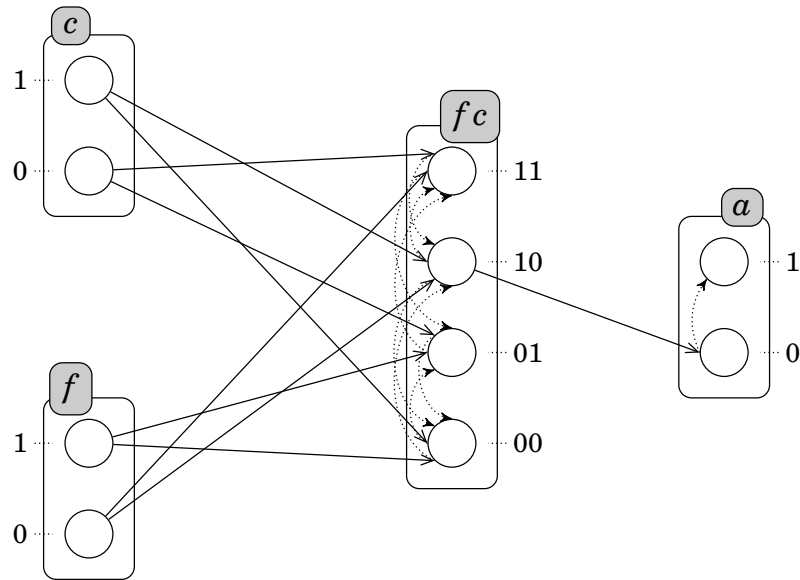


FIGURE 3.3 – Un exemple de Frappes de Processus standards avec une sorte coopérative  $fc$ . L'action  $fc_{10} \rightarrow a_0 \uparrow a_1$  modélise une coopération entre  $f_1$  et  $c_0$  pour faire bondir le processus actif de  $a$ .

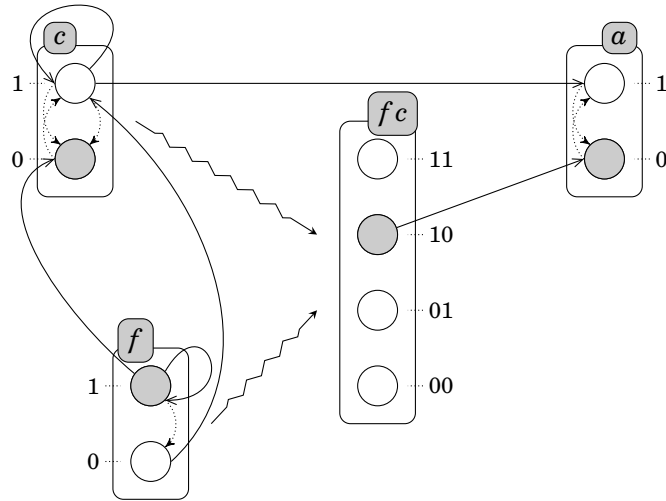


FIGURE 3.4 – Amélioration du modèle de Frappes de Processus de la figure 3.1 à l'aide de la sorte coopérative  $fc$ . Les processus de cette sorte représentent les différents sous-états formés par les deux sortes  $f$  et  $c$ . Ainsi,  $fc_{00}$  représente le fait que  $f_0$  et  $c_0$  sont actifs, etc. Les actions permettant la mise à jour de cette sorte coopérative n'ont pas été représentées explicitement mais sont symbolisées par les deux flèches en zigzag provenant de  $f$  et  $c$ .

Parallèlement à cette problématique de modélisation, il apparaît naturel de vouloir intégrer des informations de préemption entre les actions afin d'affiner la dynamique. À ce stade, on notera qu'un nombre important de formalismes ont suivi le même processus de développement, c'est-à-dire que partant du formalisme le plus large possible, des éléments de modélisation ont été ajoutés afin de restreindre la dynamique. Nous pensons aux arcs inhibiteurs (ou encore aux arcs de tests/lecture) sur les réseaux de Petri [144] ou simplement aux ensembles de priorité dans le  $\pi$ -calcul et le  $\pi$ -calcul attribué [89].

Procédant de ce raisonnement, nous avons voulu modéliser des contraintes temporelles, toute action modélisant une réaction très rapide étant par exemple systématiquement jouée avant les actions modélisant des réactions plus lentes. D'autres contraintes peuvent aussi être prises en compte, comme la concurrence entre réactions, mais aussi la représentation de comportements propres à la modélisation et n'ayant pas nécessairement de sens biologique.

Ce travail est au cœur de la thèse de doctorat de Maxime Folschette, qui a prolongé celle de Loïc Paulevé. Cette thèse s'est ainsi focalisée sur différentes formes alternatives de Frappes de Processus en intégrant les notions de préemption et de simultanéité d'une action par rapport à une autre.

La préemption permet à une action d'avoir priorité sur une autre ou, du point de vue inverse, permet d'empêcher le jeu d'une action dans une situation où elle pourrait normalement être jouable selon la dynamique des Frappes de Processus standards. Une telle préemption peut être opérée de façon généralisée comme c'est le cas pour les Frappes de Processus avec classes de priorités, où chaque action peut bloquer l'ensemble des actions de priorité inférieure ; ou de façon ponctuelle, comme au sein des Frappes de Processus avec arcs neutralisants, qui permet de définir des relations plus fines de préemption entre actions individuelles (sans que cette préemption ne soit généralisée à un ensemble d'autres actions). À l'inverse, la simultanéité entre actions permet de s'assurer qu'un ensemble de frappes est joué de façon simultanée, ou plus généralement qu'un ensemble de processus bondit en même temps, comme permettent de le représenter les Frappes de Processus avec actions plurielles.

Les apports de ces formes alternatives de Frappes de Processus permettent de restreindre la dynamique d'un modèle par rapport aux Frappes de Processus standards. Elles se posent en alternatives à l'ajout de paramètres stochastiques dans les Frappes de Processus qui permettent d'ajouter une dimension probabiliste dans ce formalisme. Leur principal atout est de renforcer la puissance d'expression des Frappes de Processus, ce qui a pour conséquence de simplifier l'écriture et la lecture des modèles, mais aussi d'offrir de nouvelles capacités de modélisation. Par ailleurs, ces différents formalismes sont tous compatibles avec les méthodes d'analyse statique développées, ce qui assure de pouvoir étudier efficacement la dynamique des modèles créés.

La définition du formalisme des Frappes de Processus avec classes de priorité a été publiée dans [68].

Nous rappelons ici le principe des Frappes de Processus avec  $k$  classes de priorité, et évoque-



rons, dans le chapitre suivant, les liens entre les différents formalismes associés.

Pour tout entier naturel  $k$  non nul, les *Frappes de Processus avec  $k$  classes de priorité* sont des Frappes de Processus dont l'ensemble des actions est partitionné en  $k$  ensembles, chacun étant associé à une classe de priorité distincte. Cela signifie qu'une action est jouable dans un état si et seulement si, en plus de la condition de la présence du frappeur et de la cible, il n'existe aucune autre action appartenant à une classe de priorité plus grande qui soit aussi jouable dans cet état. Il est à noter que les classes de priorité sont étiquetées de façon décroissante par les entiers de l'ensemble  $\llbracket 1; k \rrbracket$  en fonction de l'importance de la priorité ; autrement dit, la classe de priorité 1 contient les actions les plus prioritaires, ne pouvant jamais être préemptées, tandis que la jouabilité d'une action de la classe de priorité  $k$  ne peut pas empêcher le jeu d'une autre action. Un exemple de ce type de modèle est donné par la figure 3.5, où les différentes priorités sont signifiées par des étiquettes numérotées sur les actions.

Cette représentation basée sur des classes de priorité permet de modéliser un système dont les actions peuvent être distinguées en plusieurs classes en fonction de leur urgence, de leur vitesse d'exécution, ou encore d'autres facteurs leur donnant prévalence sur d'autres. Chaque action peut donc en préempter un ensemble d'autres en fonction de sa classe de priorité. Cela permet une représentation compacte des rapports de priorité entre actions ou, autrement dit, de leur ordonnancement, qui présente néanmoins encore quelques lacunes. Les phénomènes d'accumulation, notamment, n'y sont pas représentés ; un cycle d'actions prioritaires ne peut jamais être interrompu par une action moins prioritaire, menant à un cycle infini et pouvant contredire la réalité biologique. De plus, les classes de priorité définies pour un modèle sont invariables ; certains modèles pourraient cependant nécessiter l'évolution de certaines classes de priorité en fonction de la présence ou de l'absence d'un composant dans un état donné. Enfin, elles peuvent ne pas permettre la précision nécessaire à une représentation fidèle de certains modèles, notamment lorsqu'il est nécessaire de définir des préemptions ponctuelles comme le permettent les Frappes de Processus avec arcs neutralisants présentées dans la thèse de Maxime Folschette.

### 3.4.6 Les Frappes de Processus avec priorités

**Définition 3.8** (Frappes de Processus avec  $k$  classes de priorité). Si  $k \in \mathbb{N}^*$ , les *Frappes de Processus avec  $k$  classes de priorité* sont définies par un triplet  $\mathcal{PH} = (\Sigma; \mathcal{L}; \mathcal{H}^{(k)})$ , où  $\mathcal{H}^{(k)} = (\mathcal{H}^{(1)}; \dots; \mathcal{H}^{(k)})$  est un  $k$ -uplet, et :

- $\Sigma \stackrel{\text{def}}{=} \{a, b, \dots\}$  est l'ensemble fini et dénombrable des *sortes* ;
- $\mathcal{L} \stackrel{\text{def}}{=} \bigotimes_{a \in \Sigma} \mathcal{L}_a$  est l'ensemble fini des *états*, où  $\mathcal{L}_a = \{a_0, \dots, a_{l_a}\}$  est l'ensemble fini et dénombrable des *processus* de la sorte  $a \in \Sigma$  et  $l_a \in \mathbb{N}^*$ . Chaque processus appartient à une unique sorte :  $\forall (a_i; b_j) \in \mathcal{L}_a \times \mathcal{L}_b, a \neq b \Rightarrow a_i \neq b_j$  ;
- pour tout  $n \in \llbracket 1; k \rrbracket$ ,  $\mathcal{H}^{(n)} \stackrel{\text{def}}{=} \{a_i \rightarrow b_j \uparrow b_l \mid (a; b) \in \Sigma^2 \wedge (a_i; b_j; b_l) \in \mathcal{L}_a \times \mathcal{L}_b \times \mathcal{L}_b \wedge b_j \neq b_l \wedge a = b \Rightarrow a_i = b_j\}$  est l'ensemble fini des *actions de priorité  $n$* .

On note  $\mathcal{H} \stackrel{\text{def}}{=} \bigcup_{n \in \llbracket 1; k \rrbracket} \mathcal{H}^{(n)}$  l'ensemble de toutes les actions et, pour tout  $n \in \mathbb{N}^*$  et  $h \in \mathcal{H}^{(n)}$ ,  $\text{prio}(h) \stackrel{\text{def}}{=} n$ .

On réutilise de surcroît les notations définies précédemment concernant les états et l'extraction de la sorte d'un processus.

À l'instar de ce que nous avons vu précédemment, il faut définir un opérateur de jouabilité pour déterminer la dynamique des Frappes de Processus avec  $k$  classes de priorité. Cependant, à l'inverse de celui des Frappes de Processus standards, il faut ici prendre en compte la possible présence d'actions jouables appartenant à des classes de priorité supérieures. Pour cela, il est suffisant de vérifier que le frappeur et la cible de toute action de priorité plus importante ne sont pas simultanément présents. En effet, prenons deux actions  $g, h \in \mathcal{H}$  avec :  $\text{prio}(g) < \text{prio}(h)$ , et un état  $s \in \mathcal{L}$  tel que  $\text{frappeur}(g) \in s \wedge \text{cible}(g) \in s$  ; Deux cas de figures sont alors possibles :

- l'action  $g$  est jouable dans  $s$  — autrement dit, aucune autre action de priorité plus importante ne la préempte — et elle préempte  $h$  en conséquence,
- l'action  $g$  n'est pas jouable dans  $s$ , ce qui signifie qu'elle est préemptée par une action de priorité plus importante, qui préempte alors aussi l'action  $h$ .

Dans les deux cas,  $h$  n'est pas jouable, ce qui montre qu'il est suffisant de n'observer que la présence simultanée du frappeur et de la cible de chaque action de priorité supérieure pour déterminer la jouabilité de  $h$ . Nous obtenons alors l'opérateur de jouabilité donné à la Définition 3.9.

**Définition 3.9** (Opérateur de jouabilité ( $F_p : \mathcal{H} \rightarrow F$ )). L'opérateur de jouabilité des Frappes de Processus avec  $k$  classes de priorité est défini par :

$$\forall h \in \mathcal{H}, F_p(h) \equiv \text{frappeur}(h) \wedge \text{cible}(h) \wedge \left( \bigwedge_{\substack{g \in \mathcal{H}^{(n)} \\ n < \text{prio}(h)}} \neg (\text{frappeur}(g) \wedge \text{cible}(g)) \right)$$

*Exemple.* Nous illustrons les possibilités offertes par l'introduction des classes de priorité par la figure 3.5 qui représente un modèle de Frappes de Processus avec trois classes de priorité

$\mathcal{PH} = (\Sigma, \mathcal{L}, \mathcal{H}^{(3)})$ . Celui-ci reprend la structure du modèle de Frappes de Processus standards de la figure 3.4, en y ajoutant trois classes de priorité permettant de distinguer trois types d'actions :

- les actions de  $\mathcal{H}^{(1)}$  permettent d'assigner une priorité maximale aux actions mettant à jour la sorte coopérative  $fc$ , et peuvent être considérées comme « instantanées » du point de vue du reste du modèle ;
- les actions de  $\mathcal{H}^{(2)}$  assurent que la sorte  $a$  est mise à jour immédiatement en fonction de l'évolution de  $f$  et  $c$ , et peuvent être vues comme « urgentes » par rapport aux actions de  $\mathcal{H}^{(3)}$  ;
- enfin, les actions restantes sont par conséquent considérées comme « lentes » ou « peu urgentes » en regard du reste du modèle ; il s'agit des actions de  $\mathcal{H}^{(3)}$ , qui représentent des processus biologiques plus lents.

Comme expliqué dans la thèse de Maxime Folschette, assigner la priorité maximale aux actions permettant la mise à jour des sortes coopératives permet d'éviter les comportements indésirables décrits à la page 50. De même, accorder aux actions de  $\mathcal{H}^{(2)}$  le statut d'« urgentes » permet de s'assurer qu'elles seront jouées avant les actions de  $\mathcal{H}^{(3)}$ . Dans ce modèle, cela se traduit par le fait que l'activation ou la désactivation de  $a$  est forcée lorsque  $c$  et  $f$  évoluent, ce qui restreint la dynamique aux seuls comportements désirés. En effet, la seule dynamique possible, en partant de l'état initial  $\langle a_0, c_0, f_1, fc_{10} \rangle$ , consiste en un comportement stationnaire oscillant, où  $c$  et  $a$  sont alternativement activés et désactivés, interrompu par la désactivation de  $f$  qui entraîne irrémédiablement celle de  $c$ , sans possibilité de le ré-activer par la suite, et provoque cette fois un comportement stationnaire constant (qui se traduit en Frappes de Processus par un point fixe où  $a$  reste indéfiniment à sa dernière valeur (actif ou non). Le comportement stationnaire est donné par le scénario suivant, jouable dans l'état initial  $\langle a_0, c_0, f_1, fc_{10} \rangle$  :

$$\begin{aligned} fc_{10} \rightarrow a_0 \uparrow a_1 :: f_1 \rightarrow c_0 \uparrow c_1 :: c_1 \rightarrow fc_{10} \uparrow fc_{11} :: \\ c_1 \rightarrow a_1 \uparrow a_0 :: c_1 \rightarrow c_1 \uparrow c_0 :: c_0 \rightarrow fc_{11} \uparrow fc_{10} \end{aligned}$$

L'interruption de ce comportement stationnaire se fait grâce à l'auto-action  $f_1 \rightarrow f_1 \uparrow f_0$ , qui est de priorité 3, et donc jouable uniquement dans les deux états suivants :  $\langle a_1, c_0, f_1, fc_{10} \rangle$  et  $\langle a_0, c_1, f_1, fc_{11} \rangle$ . Depuis le premier état, la désactivation est opérée par le scénario suivant :

$$f_1 \rightarrow f_1 \uparrow f_0 :: f_0 \rightarrow fc_{10} \uparrow fc_{00} ,$$

qui termine dans l'état  $\langle a_1, c_0, f_0, fc_{00} \rangle$  et conserve donc le processus  $a_1$ , tandis que depuis le deuxième état, la désactivation est opérée par le scénario :

$$f_1 \rightarrow f_1 \uparrow f_0 :: f_0 \rightarrow fc_{11} \uparrow fc_{01} :: f_0 \rightarrow c_1 \uparrow c_0 :: c_0 \rightarrow fc_{01} \uparrow fc_{00} ,$$

qui termine en  $\langle a_0, c_0, f_0, fc_{00} \rangle$  et conserve cette fois le processus  $a_0$ .

Après avoir présenté le formalisme des Frappes de Processus et l'une de leurs extensions avec priorités, nous allons introduire, dans le chapitre suivant, les méthodes d'analyse associées.

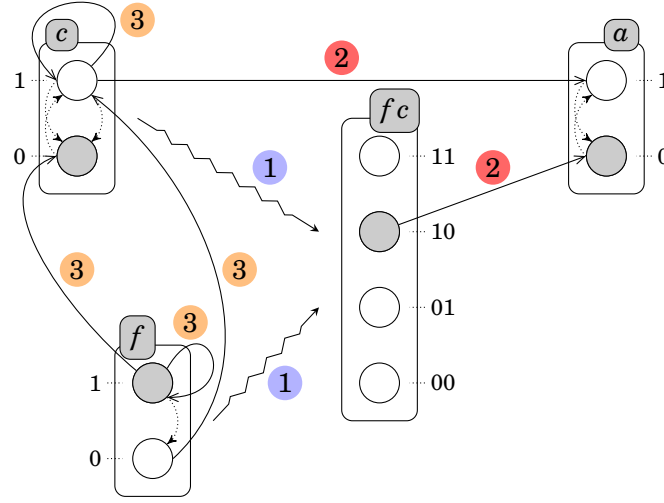


FIGURE 3.5 – Exemple de Frappes de Processus avec 3 classes de priorité. Ce modèle est issu de celui de la figure 3.4 auquel ont été rajoutées des classes de priorité. Les étiquettes numérotées (de 1 à 3) placées contre les flèches représentant les actions symbolisent leur appartenance à une classe de priorité donnée ; ainsi, on a notamment :  $\mathcal{H}^{(2)} = \{f c_{10} \rightarrow a_0 \uparrow a_1; c_1 \rightarrow a_1 \uparrow a_0\}$ .

### 3.5 Synthèse

Dans ce chapitre, nous avons défini le contexte formel de la modélisation de systèmes biologiques, plus spécifiquement des réseaux de régulation. Après avoir résumé les principaux formalismes existants et avoir rappelé leurs limites, nous avons introduit le paradigme des Frappes de Processus sur lequel nous nous concentrons depuis cinq ans. Nous avons notamment introduit le mécanisme de raffinement successif, par ajout de contraintes, de coopérations ou de priorités, permettant l'élaboration progressive de modèles signifiants. Dans le prochain chapitre, nous allons montrer comment, grâce à une représentation atomique des interactions entre composants, nous avons pu établir des méthodes efficaces d'analyse de la dynamique associée.



Dans le chapitre précédent, nous avons présenté les principaux formalismes désormais utilisés en biologie des systèmes pour représenter des mécanismes de régulation. Parce que les approches existantes souffraient à analyser des réseaux de grande taille, nous avons pris le parti d'identifier une sous-classe de certains paradigmes existants : les Frappes de Processus, et leurs extensions avec priorités. Dans ce chapitre, l'introduction de ce formalisme sera justifiée en montrant son adéquation à plusieurs questions biologiques difficiles à traiter. Nous présentons les principales idées mises en œuvre dans le cadre des Frappes de Processus afin de répondre à certaines des principales questions soulevées par les biologistes dans leurs collaborations avec les mathématiciens et informaticiens. Nous en illustrons les avantages sur des exemples concrets tirés de la biologie avec différents jeux de tests. Enfin, un positionnement global de cette contribution dans le cadre des modèles formels, notamment ceux mis en œuvre en biologie des systèmes, sera présenté.

Ces travaux ont été menés dans le cadre de la thèse de doctorat de Loïc Paulevé [133], puis approfondis au cours de la thèse de Maxime Folschette [64]. Une partie de ces recherches se sont tenues dans le cadre du projet ANR BioTempo, du projet CNRS PEPHII Circlock et du projet régional GRIOTE. Ces projets, impliquant différentes équipes de bio-informaticiens et biologistes, ont nourri nos réflexions en termes de besoins et d'objectifs applicatifs.

Les principes d'analyse statique pour les Frappes de Processus ont été présentés dans un article en revue (TCSB [139]) pour l'étude des points fixes, et dans un autre article en revue (MSCS [142]) pour l'analyse des propriétés d'accessibilité successive. La méthodologie afférente à l'analyse de ces modèles avec une sémantique stochastique a été proposée dans un troisième article en revue (IEEE TSE [140]). Une discussion au sujet de l'apport de la sémantique stochastique dans la modélisation des Frappes de Processus a été entamée dans un article en conférence

[31], puis approfondie dans un article en revue [37].

La synthèse de ce chapitre, sans s’y réduire, provient partiellement du chapitre de livre [135].

## 4.1 Quelles sont les propriétés d’intérêt ?

Nous avons choisi d’adopter une démarche de modélisation par raffinement successif. Cela signifie que, partant d’un modèle exprimant la dynamique la plus large possible, nous réduisons celle-ci progressivement en ajoutant des hypothèses sur le modèle et/ou en vérifiant la validité de certaines propriétés. Cette section présente les principales familles de propriétés biologiques qui nécessitent d’être formalisées.

### 4.1.1 Recherche de points fixes

La recherche de points fixes peut apporter des informations fondamentales dans l’étude des réseaux de régulation biologique. Un point fixe représente généralement un état stable pour le système — tout du moins du point de vue du modèle. La détection de plusieurs points fixes, par exemple, met en valeur l’existence d’un embranchement dans la dynamique du système, et le point fixe atteint dépendra du chemin suivi dans le graphe des états.

Les Frappes de Processus standards bénéficient d’outils puissants permettant la recherche de points fixes, c’est-à-dire d’états dans lesquels le modèle ne peut plus évoluer car aucune action n’est jouable. Nous rappelons ici les méthodes de recherche de points fixes développées dans [139] et qui reposent sur la recherche de  $n$ -cliques.

La recherche de points fixes dans des Frappes de Processus standards  $\mathcal{PH} = (\Sigma; \mathcal{L}; \mathcal{H})$  peut se ramener à une recherche de  $|\Sigma|$ -cliques dans un graphe particulier construit à partir de  $\mathcal{PH}$  et appelé *graphe sans-frappe*. L’idée principale de cette recherche repose sur la causalité restreinte des actions, qui implique que si les processus frappeur et cible d’une action sont tous deux actifs dans un état, alors cette action est jouable. Ainsi, les points fixes sont les actions dans lesquelles il n’existe pas de tel couple de processus actifs, et donc pas d’action jouable. La construction du graphe sans-frappe repose sur cette constatation : ses nœuds sont les processus du modèle  $\mathcal{PH}$  d’origine qui ne sont pas frappés par une auto-action, et deux nœuds ne sont reliés par une arête (non orientée) que si les deux processus correspondants ne sont *pas* reliés par une action. La division du graphe en sortes en fait un graphe  $n$ -parti avec  $n \leq |\Sigma|$  le nombre de sortes ayant au moins un processus non frappé par une auto-action. Le théorème 4.1 stipule alors que les points fixes de  $\mathcal{PH}$  sont exactement les  $|\Sigma|$ -cliques de son graphe sans-frappe, c’est-à-dire les ensemble de nœuds tous connectés entre eux, avec exactement un nœud par partie.

**Définition 4.1.** Soient  $\mathcal{PH} = (\Sigma, \mathcal{L}, \mathcal{H})$  des Frappes de Processus standards. Le graphe sans-frappe de  $\mathcal{PH}$  est le graphe non orienté  $(V, E)$  où :

$$\begin{aligned} V &\stackrel{\text{def}}{=} \bigcup_{a \in \Sigma} \{a_i \in \mathcal{L}_a \mid \forall a_k \in \mathcal{L}_a, \nexists a_i \rightarrow a_k \in \mathcal{H}\} \\ E &\stackrel{\text{def}}{=} \{\{a_i, b_j\} \in V \times V \mid a \neq b, \forall b_k \in \mathcal{L}_b, \nexists a_i \rightarrow b_k \in \mathcal{H} \\ &\quad \wedge \forall a_l \in \mathcal{L}_a, \nexists b_j \rightarrow a_l \in \mathcal{H}\} \end{aligned}$$

**Définition 4.2.** Si  $n \in \mathbb{N}$ , un graphe  $(V, E)$  est  $n$ -parti si et seulement si :

- $V = \bigcup_{k \in \llbracket 1; n \rrbracket} V_k$ ,
- $\forall k, k' \in \llbracket 1; n \rrbracket, k \neq k' \Rightarrow V_k \cap V_{k'} = \emptyset$ ,
- $\forall \{a_i, b_j\} \in E, \exists k, k' \in \llbracket 1; n \rrbracket, k \neq k' \wedge a_i \in V_k \wedge b_j \in V_{k'}$ .

**Définition 4.3.** Soit  $(V, E)$  un graphe non orienté. Un sous-ensemble  $C \subseteq V$  de ses nœuds est une  $|C|$ -clique si et seulement si :  $\forall \{a_i, b_j\} \in C \times C, \{a_i, b_j\} \in E$ .

**Théorème 4.1.** Les points fixes des Frappes de Processus standards  $\mathcal{PH} = (\Sigma, \mathcal{L}, \mathcal{H})$  sont exactement les  $|\Sigma|$ -cliques de son graphe sans-frappe.

*Exemple.* Les Frappes de Processus standards de la figure figure 3.1 page 50 contient uniquement le point fixe suivant :  $\{a_1, c_0, f_0\}$ , tandis que le modèle de la figure figure 3.4 page 57 contient les deux points fixes suivants :  $\{a_0, c_0, f_0, f c_{00}\}$  et  $\{a_1, c_0, f_0, f c_{00}\}$ .

#### 4.1.2 Interprétation abstraite (accessibilités successives)

Cette section présente brièvement et informellement l'analyse statique par interprétation abstraite développée dans [142] et inspirée des méthodes d'interprétation abstraite de programmes [47]. Celle-ci permet de répondre à des questions d'accessibilité, c'est-à-dire déterminer s'il est possible d'activer un processus depuis un état initial donné. Sa complexité polynomiale dans la taille du modèle permet une analyse très efficace de réseaux de régulation de très grande taille modélisés en Frappes de Processus.

L'étude de la dynamique des réseaux de régulation biologique se heurte généralement à la taille des modèles, et à la complexité qu'elle engendre. Les *model-checkers* traditionnels doivent en effet généralement calculer l'intégralité des comportements possibles de réseau, généralement sous la forme d'un graphe des états, afin de pouvoir produire des conclusions sur la dynamique [155]. Si une telle approche a le mérite de l'exhaustivité, elle n'est cependant pas applicable à des modèles dépassant quelques dizaines de composants, et donc totalement exclue pour tout modèle de plus d'une centaine de composants.



La forme particulière des actions des Frappes de Processus standards permet cependant de contourner en partie ces problèmes de complexité en approchant la dynamique des modèles. L'analyse statique développée dans [142] utilise en effet cette propriété pour résoudre efficacement des problèmes d'accessibilité, c'est-à-dire des problèmes qui s'énoncent de la façon suivante : « Est-il possible d'activer un processus depuis un état donné en jouant un certain nombre d'actions ? » Ou, autrement dit : « Étant donné un état  $s$  et un processus  $a_i$ , existe-t-il un scénario  $\delta \in \mathbf{Sce}(s)$  dans cet état tel que  $a_i \in s \cdot \delta$  ? »

La méthode développée pour répondre à ce type de question est basée sur deux constatations :

- on peut vérifier localement que le processus actif d'une sorte peut bondir d'un niveau vers un autre en n'observant que les actions frappant cette sorte ;
- le jeu d'une action est conditionné par au plus un processus d'une sorte autre que la cible de cette action.

De la première constatation découle le fait que l'accessibilité d'un processus peut être résolue localement, en observant les actions frappant une sorte donnée, ce qui simplifie la recherche. Une fois ce problème résolu localement pour une sorte, la deuxième constatation permet de déplacer ce problème à d'autres sortes dans le but d'activer les processus requis pour jouer chaque action nécessaire (c'est-à-dire les frappeurs de ces actions).

Il existe donc un lien de causalité récursif entre l'activation locale d'un processus, les actions requises pour le faire, et de nouveau l'activation des processus nécessaires pour jouer ces actions, etc. Afin de réduire la complexité de la méthode, une approximation de la dynamique est effectuée à chaque étape : pour chaque résolution d'accessibilité locale au sein d'une sorte, un ensemble de processus requis appartenant à d'autres sortes est produit, mais l'ordre dans lequel ces processus sont nécessaires est abstrait. Cela permet donc de déplacer l'accessibilité locale d'un processus à plusieurs accessibilités indépendantes dans d'autres sortes.

Cette approche permet de décliner une approximation supérieure (sur-approximation) et une approximation inférieure (sous-approximation) de l'ensemble de toutes les dynamiques possibles du modèle. La sur-approximation consiste à ne pas s'intéresser à l'ordre dans lequel les processus requis sont activables — et donc à l'ordre dans lequel les actions résolvant l'accessibilité locale sont jouables. Cela autorise effectivement davantage de comportements, car en pratique un processus peut ne pas être activable après certaines actions. La sous-approximation, à l'inverse, stipule que tous les processus requis doivent être activables dans tous les ordres possibles, bien qu'en pratique, tous les ordres ne soient pas intéressants pour la résolution.

Chacune de ces approximations est représentée à l'aide d'un *graphe de causalité locale*, qui est unique à chaque problème d'accessibilité, et qui formalise les liens de causalité évoqués précédemment. Enfin, nous avons donné dans [142] une propriété qui, sous certaines conditions dépendant du graphe de causalité locale correspondant à la sous-approximation, stipule que le processus donné est atteignable depuis l'état donné ; de même, une propriété complémentaire est donnée pour le graphe de causalité locale correspondant à la sur-approximation, qui permet

d'obtenir la conclusion inverse. Si aucune des deux propriétés n'est vraie, la méthode est dite non conclusive, et il est nécessaire de raffiner le problème ou le modèle.

Le calcul des deux graphes de causalité locale est polynomial dans la taille des Frappes de Processus standards sur lesquelles la méthode est appliquée, et la vérification des deux propriétés l'est dans la taille des graphes obtenus. Ainsi, cette méthode est plus efficace que les approches par force brute car elle évite l'explosion combinatoire propre à l'analyse de la dynamique. Son implémentation produit des résultats en quelques dixièmes de seconde sur des modèles de plusieurs centaines de composants, et s'avère toujours conclusive sur les exemples que nous avons étudiés.

Une partie de cette analyse statique a par la suite été exploitée dans le but d'approximer efficacement des ensembles de coupes, c'est-à-dire des ensembles de processus nécessaires à une certaine accessibilité [134]. Son utilisation dans ce cadre s'est avérée efficace sur des modèles de plusieurs milliers de composants.

### 4.1.3 Simulation stochastique

Nous avons également proposé un enrichissement des Frappes de Processus standards à l'aide de paramètres stochastiques [31, 139], l'objectif étant d'intégrer des données temporelles continues dans les modèles. Cet enrichissement est directement inspiré du  $\pi$ -calcul stochastique [148] et, plus généralement, des travaux portant sur l'analyse des réseaux de régulation génétiques via des modèles stochastiques (réseaux de Petri stochastiques [81], modèles  $\kappa$  [48] ou Biocham [61] équipés d'une sémantique stochastique, ...). Cependant, la loi exponentielle utilisée pour la loi stochastique étant jugée trop grossière, l'approche a été raffinée par l'introduction d'un paramètre supplémentaire permettant de réduire l'intervalle de tir [140].

L'introduction de données dans les Frappes de Processus standards consiste à affecter un couple de paramètres stochastiques  $(r; sa) \in \mathbb{N} \times \mathbb{R}$  à chaque action. La probabilité de tirer une action à un instant donné (sur un axe de temps continu) suit alors une loi d'Erlang en fonction de ces deux paramètres, c'est-à-dire une somme de lois exponentielles. Le premier paramètre, appelé *taux*, indique le nombre de fois qu'une action peut être tirée par unité de temps. Le second paramètre est l'*absorption de stochasticité*, qui détermine le nombre de lois exponentielles sommées pour obtenir la distribution finale.

Tout couple de paramètres stochastiques  $(r; sa)$  correspond à un intervalle de tir  $[d; D]$ , où  $d, D \in \mathbb{R}$ , pour un niveau de confiance  $\alpha \in [0; 1]$  donné, et inversement, qui peut être approximé. Cette conversion permet de raisonner sur des intervalles de tirs plutôt qu'en termes de loi d'Erlang, ce qui permet notamment de définir des fenêtres de tir pour chaque action — une action devant être tirée dans sa fenêtre avec un niveau de confiance de  $\alpha$ . Au niveau de l'intervalle de tir, les deux paramètres stochastiques ont un rôle particulier :

- Le taux détermine la distance de l'intervalle de tir par rapport à l'origine de l'axe du temps ; autrement dit, une action avec un taux élevé sera tirée plus rapidement après

sensibilisation.

- Le facteur d'absorption de stochasticité détermine la taille de l'intervalle de tir. Ainsi, augmenter ce facteur réduit la quantité  $D - d$ .

Enfin, il est aussi possible d'assigner un taux *infini* à une action ( $r = \infty$ ), ce qui a pour conséquence de rendre le tir de l'action instantané : l'action doit être jouée dès sa sensibilisation. Si plusieurs actions de taux infini sont sensibilisées en même temps, le non-déterminisme est la règle et elles peuvent donc être jouées dans n'importe quel ordre. Une action de taux infini ne possède pas de facteur d'absorption de stochasticité, car ce deuxième paramètre ne changerait rien à sa condition de tir.

*Exemple.* Définissons les fenêtres de tir données sur la figure 4.1 page 70, inspirées de [139]. Nous pouvons enrichir les Frappes de Processus standards décrites par la figure 3.4 sur la page 57 à l'aide de ces paramètres, de la façon donnée à la figure 4.2. Cette affectation nous permet notamment d'obtenir un comportement oscillant pour les sortes  $a$  et  $c$ , et une auto-désactivation de  $f$  après plusieurs oscillations. Les paramètres du couple  $\mathbf{c}$  ont en effet été conçus de façon à ce que l'arrêt de l'horloge, provoqué par la désactivation de  $f$ , survienne après la formation de 7 rayures.

Les modèles de Frappes de Processus standards comportant des paramètres stochastiques peuvent être simulés afin d'observer un exemple d'exécution du modèle étudié. Du fait de la nature non-déterministe et probabiliste du modèle, plusieurs dynamiques sont possibles, et chaque simulation donne des résultats différents en termes de chemin d'exécution — ou, pour un même chemin d'exécution, en termes de temps de tir des actions. Il est donc possible d'effectuer un grand nombre de simulations pour obtenir des statistiques sur le comportement d'un modèle. Il est aussi possible de recourir à un *model-checker* probabiliste comme PRISM [100], qui fournit des résultats formels sur la dynamique d'un modèle, et des probabilités d'accessibilité. Cependant, une telle approche est très coûteuse en termes de temps de calcul et d'espace mémoire, ce qui empêche l'analyse de réseaux de grande taille.

## 4.2 Aperçu des résultats obtenus

### 4.2.1 L'outil Pint

Pint<sup>1</sup> est un ensemble de programmes exécutables en lignes de commande qui mettent en œuvre les différentes méthodes d'analyse des Frappes de Processus, y compris les extensions des méthodes précédentes développées pour les Frappes de Processus à priorités. Il a été initié par Loïc Paulevé pour illustrer les résultats obtenus pendant sa thèse. Loïc continue d'améliorer l'implantation, et les doctorants qui ont pris sa suite - Maxime Folschette, Emna Ben Abdallah

---

1. <http://loicpauleve.name/pint>

notamment - intègrent aussi leurs résultats à cet ensemble logiciel. Pint est distribué sous licence libre CeCILL. Une interface graphique est également disponible <sup>2</sup>.

Pint prend en entrée un langage textuel pour décrire les modèles de Frappe de Processus. Une Frappe de Processus peut être décrite par la liste de ses actions, mais aussi grâce à des macros qui facilitent l'écriture du modèle, en particulier en prenant en entrée des réseaux de régulation biologiques. La figure 4.3 présente le fichier source du modèle sous forme de Frappes de Processus raffinées, correspondant à la boucle incohérente de rétro-action décrite dans [118]. Ce fichier utilise deux macros : BRN qui calcule la dynamique généralisée d'un graphe d'interactions donné ; et COOPERATIVITY qui assure le raffinement selon les coopérations entre composants. La liste des actions peut être obtenue en utilisant l'utilitaire phc (`phc -l dump -i source.ph`).

Parmi les fonctionnalités implantées dans Pint, nous pouvons citer :

- la simulation stochastique qui prend en charge l'absorption de stochasticité (simulation non-markovienne) ;
- une implantation permettant l'obtention des points fixes du modèle ;
- un analyseur statique pour les propriétés d'accessibilité et de calcul d'ensembles de coupes (ou "cut sets") pour l'accessibilité, qui sont des ensembles d'états locaux de processus qui, s'ils sont désactivés, rendent impossible l'accessibilité étudiée [136] ;
- l'inférence de réseaux booléens/réseaux multi-valués compatibles avec le modèle de Frappe de Processus en entrée ;
- l'import/export vers différents formats utilisés pour la représentation des réseaux biologiques.

La documentation à propos de ces fonctionnalités peut être trouvée sur le site de Pint mentionné ci-dessus.

Pour finir, il convient de signaler que Pint a été intégré au portail Mobyle développé par Julien Gras dans le cadre de l'ANR BioTempo. Cette interface web <sup>3</sup> rassemble plusieurs outils logiciels utiles pour l'étude des mécanismes temporels en jeu dans les réseaux de régulation biologiques. Il est possible de définir ses propres "workflows" et de les exécuter directement sur l'infrastructure sur laquelle le portail est hébergé.

## 4.2.2 Exemples biologiques

Maintenant que nous avons établi comment construire, définir et analyser un modèle de Frappes de Processus, attaquons-nous à un exemple biologique plus réaliste pour illustrer comment le framework peut être utilisé en pratique. En particulier, nous allons détailler une étude du récepteur EGF sur un modèle à moyenne échelle pour souligner la complémentarité entre notre approche de modélisation par le raffinement d'une dynamique généralisée et nos

---

2. <http://loicpauleve.name/pint/gui>

3. <http://mobyle.biotempo.univ-nantes.fr/cgi-bin/portal.py>

méthodes statiques pour l'accessibilité. Enfin, nous présentons brièvement les résultats de Pint pour l'analyse d'accessibilité sur des réseaux à grande échelle, jusqu'à 100 composants en interaction.

#### 4.2.2.1 Étude des dynamiques de récepteur EGF

Dans cette section, nous nous intéressons à l'étude d'un modèle de récepteur de facteur de croissance épidermique.

ErbB représente la famille génétique de quatre tyrosine kinases de récepteur structurellement similaires, parmi lesquels le récepteur du facteur de croissance épidermique (EGFR ou ErbB-1) est largement étudié dans le domaine de l'oncologie. Dans les cellules saines, la production de ErbB conduit à une voie de signalisation qui régule les cellules des phases G1 à S, un point de contrôle qui détermine si une cellule doit se diviser, retarder la division ou entrer dans un état de repos. La sur-expression de ErbB est associée à de nombreux types de cancer, et les médicaments qui les ciblent, eux et leurs récepteurs, sont des traitements communs du cancer du sein, du poumon et du côlon. Dans le cas du cancer du sein, la résistance aux médicaments se manifeste chez environ 30% des patients, conduisant à un appel à d'autres cibles le long de la même voie moléculaire. Les thérapies innovantes qui exploitent la structure du réseau de régulation ont le potentiel pour faire avancer les interventions médicales sur un plan clinique. Ici, nous commençons notre investigation en considérant un modèle simplifié de vingt sortes qui ne contiennent pas toutes les sous-réactions, mais capturent la plupart des dynamiques importantes du processus de signalisation ErbB. Le graphe orienté pour ce réseau a été construit à partir de [157]. Ce modèle à moyenne échelle contient deux principales protéines intéressantes :

- Une protéine nommée EGF peut être considérée comme un composant d'entrée car elle ne comporte aucun régulateur,
- Une protéine baptisée pRB peut être considérée comme la seule sortie car une chaîne de réactions permet l'activation de cette protéine pRB, qui est responsable de la régulation de la division cellulaire.

Afin d'illustrer le raisonnement sur la dynamique du modèle en utilisant les Frappes de Processus et l'approche de raffinement de la dynamique généralisée du Graphe d'Interaction, nous appliquons nos analyses sur trois modèles différents de Frappe de Processus, avec différents niveaux de précision dans la modélisation des coopérations entre composants :

- Le modèle (1) représente la dynamique généralisée du graphe des interactions, c'est-à-dire sans aucune information concernant les règles booléennes des coopérations, et ne contient donc aucune sorte coopérative.
- Le modèle (2) est partiellement raffiné : les dynamiques sont raffinées par l'ajout de 14 sortes coopératives. En nous basant sur la littérature dans le domaine, nous avons sélectionné les composants les plus importants dans la chaîne de réactions. En effet, les

expériences de knockdown [157] montrent que l'élimination de ces composants conduit à une diminution significative de la production de la protéine pRB de sortie. Par conséquent, toutes les fonctions booléennes impliquant ces éléments ont été incluses dans ce modèle via des sortes coopératives.

- Enfin, le modèle (3) est le modèle totalement raffiné contenant toutes les coopérations, via les fonctions booléennes fournies dans [157]. Ces fonctions prennent la forme de 22 sortes coopératives qui permettent de retrouver le comportement attendu du système.

Le tableau 4.1 résume plusieurs résultats d'analyse statique sur ces modèles, basés sur les méthodes développées dans la section 4.1.2 et commentés brièvement ci-dessous.

Modèle	Points fixes	$\mathcal{P}_{\text{EGF}_1}$	$\mathcal{P}_{\text{EGF}_0}$
(1)	0	Vrai	Vrai
(2)	0	Vrai	Faux
(3)	3	Vrai	Faux

TABLE 4.1 – Résultats pour plusieurs analyses sur les trois modèles de régulation EGFR / ErbB avec 20 composants. La première colonne liste les modèles utilisés. La deuxième colonne indique le nombre de points fixes énumérés par l'analyse statique de la section 4.1.1. La troisième colonne donne le résultat de l'analyse statique décrite dans la section 4.1.2 pour la propriété  $\text{reach}(\zeta, \text{pRB}_0 \mapsto^* \text{pRB}_1)$ , où  $\zeta$  est l'état initial où tous les composants sont au niveau 0 et la protéine EGF en entrée est au niveau 1. Cette expérience a été répétée lorsque EGF est au niveau 0 et est listée dans la quatrième colonne.

*Points fixes.* Le modèle (3) est le seul modèle qui contient des points fixes. Parmi les trois états stables trouvés, deux états correspondent à une propagation totale du signal d'entrée, dans le cas où EGF est actif et dans le cas où il ne l'est pas. Les deux autres modèles ne contiennent pas de points fixes parce que certaines coopérations ne sont pas entièrement définies, conduisant à des oscillations à la suite d'un comportement non déterministe.

**Notation pour les propriétés d'accessibilité.** Dans ce qui suit, un *objectif* est un couple de processus qui décrivent l'activation d'un processus depuis un autre processus. Par exemple,  $d_0 \mapsto^* d_2$  décrit l'accessibilité d'un processus  $d_2$  depuis un état dans lequel  $d_0$  est actif. Nous appelons *propriété d'accessibilité* une propriété de la forme  $\mathcal{P} = \text{reach}(\zeta, \omega)$  où  $\zeta$  est un *contexte* initial (c'est-à-dire la généralisation d'un état) et  $\omega = P_1 :: P_2 :: \dots :: P_{|\omega|}$  est une *séquence d'objectifs*. Les objectifs sont utiles pour décrire une propriété d'accessibilité, étant donné que l'accessibilité d'un processus  $a_i$  depuis un état  $s$  se note par l'objectif :  $s[a] \mapsto^* a_i$ .

**Accessibilité de la sortie.** La principale propriété d'accessibilité intéressante ici est  $\text{pRB}_0 \mapsto^* \text{pRB}_1$ , étant donné que pRB est la seule sortie des trois modèles. Si nous supposons que tous les composants sont d'abord inactifs, il y a deux propriétés d'accessibilité intéressantes :

- $\mathcal{P}_{\text{EGF}_1} = \text{reach}(\zeta \sqcap \{\text{EGF}_1\}, \text{pRB}_0 \mapsto^* \text{pRB}_1)$ ,

$$— \mathcal{P}_{\text{EGF}_0} = \text{reach}(\zeta, \text{pRB}_0 \xrightarrow{*} \text{pRB}_1),$$

où  $\zeta$  est le contexte initial où toutes les sortes sont au niveau 0. Autrement dit,  $\mathcal{P}_{\text{EGF}_1}$  signifie que pRB peut être activée dans des conditions normales (lorsque EGF en entrée est activé), et  $\mathcal{P}_{\text{EGF}_0}$  signifie que pRB peut être activé en dépit de EGF étant à 0, un état qui peut être utilisé pour invalider un modèle défectueux. Nous notons que  $\mathcal{P}_{\text{EGF}_1}$  est, bien sûr, vrai pour tous les modèles, et  $\mathcal{P}_{\text{EGF}_0}$  est seulement vrai que pour le modèle (1) : les dynamiques généralisées sont trop permissives pour ce système. Nous notons également que le fait que  $\mathcal{P}_{\text{EGF}_0}$  est faux pour le modèle (2) est suffisant pour prouver que l'accessibilité est également impossible pour le modèle (3), en raison de la relation d'abstraction entre les deux modèles.

Cette étude met l'accent sur l'avantage d'utiliser des raffinements (partiels) de la dynamique généralisée du réseau via les Frappes de Processus : même sans une connaissance complète des fonctions logiques précises entre régulateurs, il est tout de même possible de dériver des propriétés dynamiques formelles sur le système et d'en tirer des conclusions sur n'importe quel modèle raffiné en résultant.

#### 4.2.2.2 Performances sur des réseaux à grande échelle

Nous avons mis en évidence précédemment, en section 4.1.2, que l'analyse statique des modèles de Frappes de Processus utilisant l'interprétation abstraite présente une faible complexité par rapport au model-checking exact. Le but d'une telle approche est de parvenir à des informations sur le modèle même quand les méthodes classiques de model-checking échouent, en raison de la taille des modèles. Cependant, nos méthodes peuvent ne pas donner de résultats concluants (les conditions nécessaires sont remplies, mais pas les conditions suffisantes). Le tableau 4.2 en page 75 énumère quelques-uns des temps d'exécution obtenus sur différents modèles de Frappes de Processus de voies de signalisation : un modèle du récepteur d'EGF de moyenne taille (20 éléments) et de grande taille (104 composants) [157, 158], et un modèle de récepteur des cellules T de moyenne taille (40 éléments) et de grande taille (94 composants) [94, 156]. Le nombre de sortes est en général supérieur au nombre de composants car certaines sortes sont utilisées pour coder des fonctions logiques (sortes coopératives). Ces modèles de Frappes de Processus sont disponibles sur le site de Pint mentionné précédemment. Pour chacun de ces modèles, une large gamme de propriétés d'accessibilité ont été vérifiés à partir de nombreuses conditions initiales. La vérification a été menée en utilisant les model-checkers symboliques exacts Biocham [62] et libddd [79], en plus de Pint. Pint rend traitable l'analyse formelle de la dynamique de ces modèles (grâce à la faible complexité de notre analyse statique) car, dans les cas étudiés ici, *toutes les analyses sont concluantes*.

Les travaux ultérieurs de Loïc Paulevé ont permis de mettre en évidence les ensembles de coupes ("cut-sets"), c'est-à-dire l'ensemble minimal des processus à bloquer pour retirer une accessibilité. Les résultats détaillés de cette étude complémentaire ont été présentés dans [135, 136].

	$\infty$	<b>a</b>	<b>b</b>	<b>c</b>
$[d;D]$	$[0;0]$	$[0,742;1,29]$	$[1,48;2,59]$	$[23,9;35,4]$
$(r;sa)$	$r = \infty$	$(1;50)$	$(0,5;50)$	$(0,034;100)$



FIGURE 4.1 – Exemples d'intervalles de tir et de paramètres stochastiques.

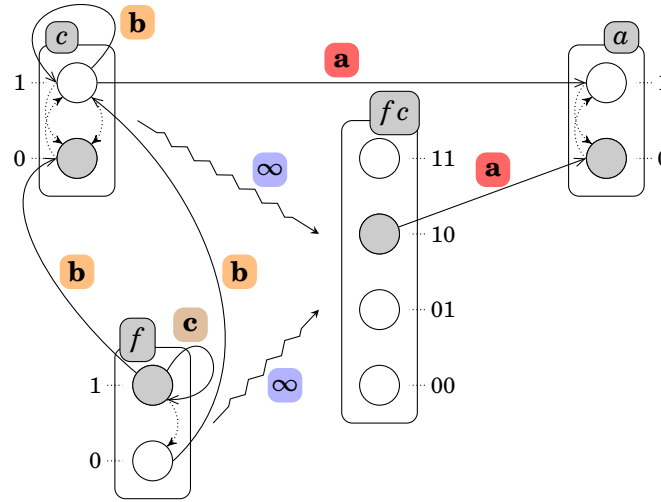


FIGURE 4.2 – Enrichissement des Frappes de Processus standards de la figure 3.4 à l'aide des paramètres stochastiques proposés à la figure 4.1. Chaque action est affectée à un couple de paramètres stochastiques repéré par une étiquette ( $\infty$ , **a**, **b**, **c**) placée contre l'action. Les ensembles d'actions mettant à jour la sorte coopérative  $fc$ , représentées par des flèches en zigzag, sont toutes affectées, possèdent toutes un taux infini ( $r = \infty$ ).

Modèle	Nb sorts	Nb procs	Nb actions	Nb états	Biocham	libddd	Pint
egfr20	35	196	670	$2^{64}$	[3s-out]	[1s-150s]	0.007s
tcrsig40	54	156	301	$2^{73}$	[1s-out]	[0.6s-out]	0.004s
tcrsig94	133	448	1124	$2^{194}$	out	out	0.030s
egfr104	193	748	2356	$2^{320}$	out	out	0.050s

TABLE 4.2 – Comparaison des temps d'exécution de différentes propriétés d'accessibilité de plusieurs modèles entre Biocham (en utilisant NuSMV), libddd [79], et Pint. Un dépassement des capacités mémoires de la machine est noté "out". Comme les temps d'exécution varient considérablement selon les propriétés d'accessibilité, nous donnons les durées minimales et maximales.



```
(* declaration of sorts *)
process a 1
process b 1
process c 1

(* generalized dynamics *)
BRN([a 1 -> + b; a 1 -> - c; b 1 -> + c])

(* cooperativity: c becomes 1 only when a=0 and b=1 *)
COOPERATIVITY([a;b] -> c 0 1, [[0;1]])
```

FIGURE 4.3 – Fichier source pour le modèle de boucle incohérente de rétroaction décrite dans [118].

---

À notre connaissance, Pint est actuellement le seul outil capable d’effectuer une accessibilité (transitoire) pour les dynamiques de réseaux à une telle échelle, allant de quelques centaines à plusieurs milliers de composants.

## 4.3 Positionnement des Frappes de Processus par rapport aux formalismes existants

### 4.3.1 Équivalences avec les réseaux booléens et le modèle de Thomas

Nous nous sommes également intéressés aux liens formels entre les différentes sémantiques de Frappes de Processus et les autres modélisations classiques pour la représentation des réseaux de régulation biologique. Notamment le modèle de Thomas et les réseaux booléens, deux formalismes proches développés spécifiquement pour la représentation de réseaux de régulation. Dans [68], nous avons montré qu’il est possible de représenter ces deux formalismes de façon exacte grâce à l’ajout de priorités dans les Frappes de Processus [68].

### 4.3.2 Équivalences entre les Frappes de Processus

Les méthodes d’analyse statique présentées dans ce chapitre s’appliquent plus largement aux Frappes de Processus canoniques, une classe particulière introduite dans la thèse de Maxime Folschette [64]. Comme cela a été démontré, cette classe est aussi expressive que tous les formalismes de Frappes de Processus à priorités développés jusque là [68]. Les traductions associées ont été exhibées dans le manuscrit de thèse de Maxime Folschette, ce qui assure notamment qu’il est toujours possible de naviguer entre les formalismes, afin de profiter de leurs avantages respectifs, et notamment de l’utilisation des analyses statiques mentionnées ci-dessus à tous les autres types de Frappes de Processus, moyennant une traduction du modèle (voir figure 4.4).

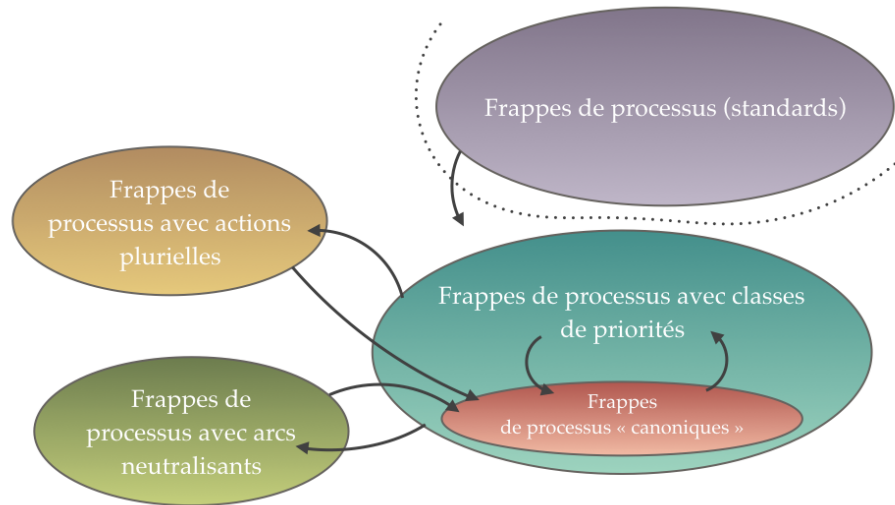


FIGURE 4.4 – Synthèse des équivalences (en termes de (bi)simulation faible) entre les différentes Frappes de Processus.

### 4.3.3 Équivalences avec les réseaux de Petri

De par leur structure, les Frappes de Processus évoquent intuitivement une classe particulière de réseaux de Petri. Pour fixer plus formellement cette intuition, nous nous sommes intéressés à deux types de traduction différentes des Frappes de Processus vers les réseaux de Petri.

Cette traduction peut également s'inscrire dans une volonté plus générale d'utiliser ce formalisme pour représenter des processus de régulation biologique [35]. De plus, cela peut permettre d'appliquer des méthodes de dépliage [10] afin de réduire le parallélisme dans les modèles et faciliter en partie l'analyse par *model-checking*.

La première traduction, présentée dans [141], ne considère que les Frappes de Processus standard. Elle fait correspondre une place à chaque processus, et produit donc un réseau de Petri sauf et n'utilisant pas d'arcs inhibiteurs.

La seconde traduction, proposée dans [64], s'applique à toutes les sémantiques de Frappes de Processus (avec actions plurielles, arcs neutralisants ou classes de priorités, mais aussi standards). Elle associe à toute sorte une place dans le réseau de Petri produit. Les jetons de cette place représentent alors le processus actif dans la sorte correspondante. Les actions sont, quant à elles, représentées par une transition comportant les arcs ordinaires, de lecture et inhibiteurs nécessaires pour reproduire la dynamique recherchée.

Cela nous a notamment permis de montrer que les Frappes de Processus sont expressivement incluses dans les réseaux de Petri. De plus, l'avantage de cette deuxième traduction proposée est de ne représenter qu'une place par sorte, d'être bornée et d'utiliser le système d'accumulation des jetons pour représenter les processus, même dans un cas multi-valué (non-booléen).

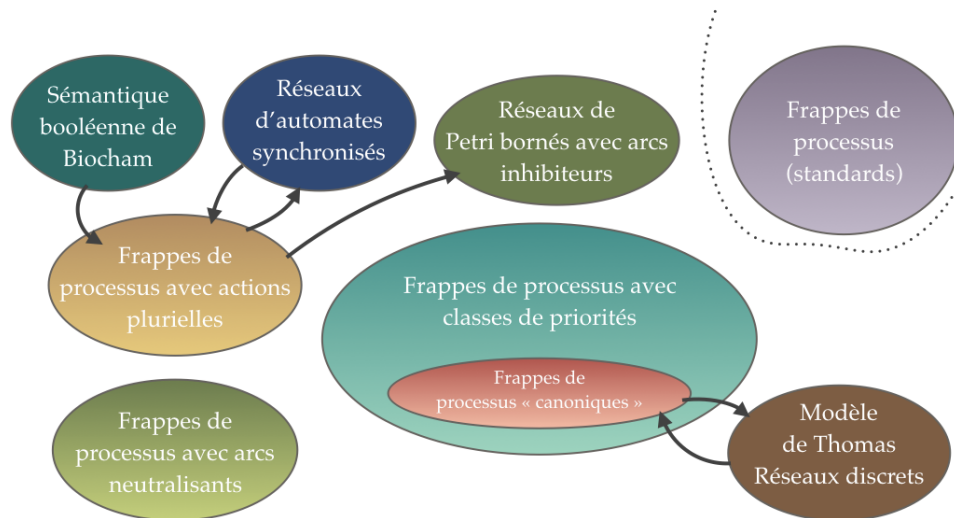


FIGURE 4.5 – Synthèse des équivalences (en termes de (bi)simulation faible) entre les Frappes de Processus et quelques-uns des principaux formalismes de la littérature.

#### 4.3.4 Équivalences complémentaires

Nous avons également étudié d'autres traductions, partant de modèles existants vers les Frappes de Processus. Au-delà de la traduction naturelle des réseaux de régulation biologiques discrets vers les Frappes de Processus, nous avons également considéré [64] une traduction depuis la sémantique booléenne de Biocham [62]. La figure 4.5 propose une synthèse des liens ainsi mis en évidence.

### 4.4 Synthèse

Nous avons résumé dans ce chapitre les méthodes d'analyse qualitative que nous avons mises en place pour les réseaux de régulation biologiques. L'un des objectifs majeurs de notre travail est de traiter de très grands réseaux : ce faisant, il devient alors possible de travailler sur des modèles plus réalistes qui sont d'un grand intérêt pour les biologistes.

Le but ultime de ce paradigme de modélisation est de nous permettre d'analyser le comportement opérationnel des systèmes biologiques, pour comprendre non seulement ce qui peut éventuellement se produire et ce qui se produira inéluctablement, mais aussi de comprendre comment empêcher certains événements de se produire, car ils peuvent amener sur des états pathologiques.

Les régulations (qu'il s'agisse d'activations ou d'inhibitions) sont généralement caractérisées par le franchissement d'un certain niveau de seuil de concentration ou de niveau d'expression. Bien que chacune de ces réactions soit, seule, très simple, la complexité globale du modèle augmente avec le nombre d'agents en interaction et le nombre d'états dans lesquels ils peuvent se

trouver. Le nombre de combinaisons possibles de ces états conduit à une croissance exponentielle de la taille du modèle résultant. Malheureusement, la taille importante des modèles nous empêche habituellement de les analyser en utilisant des approches classiques. Il est impératif de concevoir des méthodes qui permettent d'abstraire intelligemment l'information contenue dans les modèles. C'est ce travail auquel nous nous sommes attelés. Le modèle des Frappes de Processus repose sur une décomposition discrète (et atomique) des interactions entre les composants biologiques.

Initialement inspirées par le  $\pi$ -calculus, les Frappes de Processus reposent sur des principes simples :

- Plutôt que de se concentrer sur le concept d'état global, notre méthode commence par rassembler dans une entité fondamentale (que nous avons nommée *sorte*) les différents états possibles de chaque composant du système. Cette représentation repose sur l'idée que, à tout moment, un composant ne peut être que dans un seul de ces états (nommé *processus* courant). Ce principe nous permet d'éviter d'énumérer directement l'espace d'états. Il s'agit là d'une propriété très souhaitable lorsque l'espace est d'une dimension importante et qu'une grande partie des états globaux sont inaccessibles ou inexprimés.
- Pour étudier le fonctionnement du système, nous partons d'un état initial (un tuple composé d'un processus de chaque sorte). Nous nous intéressons alors à tous les états dans lesquels un composant peut être après une interaction entre deux des composants présents dans le tuple. Nous ne prenons en compte que ces deux composants, et seul celui qui subit une réaction change alors d'état. Les informations devant être capturées pendant ces transitions d'état sont ainsi réduites à ce que l'on nomme une frappe de processus. La liste complète des frappes nous permet de reconstituer facilement chaque scénario du système. De plus, il est possible qu'après certaines frappes, certains événements ne soient plus possibles. Cela indique que certaines relations de précédence doivent être, à certains moments, réintroduites dans l'ensemble des données à traiter.

Les Frappes de Processus sont bien adaptées à une analyse par raffinement successif du modèle. C'est l'une des grandes forces de ce framework.

Nous avons mis en évidence des méthodes d'analyse statique bien adaptées pour l'étude des comportements discrets ou multi-niveaux. Nous avons également commencé à explorer des sémantiques stochastiques.

Les résultats que nous avons obtenus sur l'analyse des Frappes de Processus nous ont motivé à avancer et à examiner d'autres perspectives. Parmi celles-ci figure l'apprentissage automatique de réseaux de régulation biologique, un travail que nous avons entamé et que nous allons présenter dans le chapitre suivant. Cette étude repose notamment sur l'utilisation de Answer Set Programming (ASP). Dans [66], nous avons ainsi établi un lien entre un modèle de Frappe de Processus et le Graphe d'Interaction correspondant, lien étendu ensuite à la paramétrisation des modèles représentés selon l'approche de René Thomas [168]. Ceci a été réalisé en mettant au point une méthode d'énumération des paramétrages efficacement compatibles.



## INFÉRER, VOIRE APPRENDRE (AUTOMATIQUEMENT)

Dans les chapitres précédents, nous avons mis en évidence l'enjeu que représente la modélisation de la dynamique de systèmes biologiques. Nous avons notamment souligné notre contribution, visant à identifier des sous-classes intéressantes de modèles sur lesquelles des méthodes efficaces d'analyse ont pu être développées. Ces analyses, aussi précieuses soient-elles dans une démarche de co-construction d'un modèle avec des équipes de biologistes, se heurtent à une limite majeure : elles sont pensées comme une aide à la construction. Elles ne font pas un lien direct avec la quantité de données que les biologistes produisent désormais, aidés par des techniques de séquençage à haut-débit. C'est dans l'optique de s'adapter à cette tendance et d'aider à automatiser la démarche de modélisation que se situe ce chapitre. Là où nous nous sommes focalisés, dans le chapitre précédent, sur des approches de model-checking, nous nous penchons ici sur des approches logiques, qui nous ont ouvert la voie à des techniques d'inférence automatique au sein de nos modèles. Après avoir résumé l'apport de ces approches - et notamment de la programmation ASP - dans l'inférence de paramètres au sein de réseaux de régulation biologique, nous présentons quelques travaux et réflexions récentes autour de la reconstruction d'un réseau de régulation à partir de données d'exécution. Nous nous plaçons en cela dans une perspective de biologie exécutable, ce qui nous permettra de mettre en valeur les promesses d'une combinaison d'approches, inspirées à la fois du model-checking, et de l'apprentissage automatique, pour la construction et le traitement de réseaux de régulation.

Ces travaux ont été menés notamment dans le cadre de la thèse de doctorat de Maxime Folschette, ainsi qu'au cours de mes séjours au National Institute of Informatics depuis 2013, en collaboration avec Katsumi Inoue et Tony Ribeiro.

L'inférence des paramètres discrets et du graphe d'interaction des réseaux de régulation biologiques via l'utilisation conjointe des Frappes de Processus et d'ASP a fait l'objet d'un

article en revue [67], d'une publication en conférence internationale [66] et d'une publication en workshop [65]. L'apprentissage automatique de programmes logiques à partir de données de séries temporelles a été présenté dans un article en revue [151] et une publication courte en conférence internationale [150].

## **5.1 Approches logiques : induction, déduction, abduction et machine learning**

La collaboration avec l'équipe de Katsumi Inoue au National Institute of Informatics, entamée en 2011, m'a donné l'opportunité d'investir un thème de recherche connexe, autour de l'apport de la programmation logique pour l'apprentissage automatique de réseaux de régulation. Ces travaux, qui recouvrent différents thèmes intimement liés, sont en cours à l'heure où ce manuscrit est rédigé, mais de premiers résultats significatifs ont été obtenus. Dans cette section, je résume les enjeux et obstacles liés à cette problématique, ainsi que la démarche que nous avons commencée à développer pour atteindre ces objectifs.

## **5.2 Inférence de paramètres discrets des réseaux de régulation biologiques**

Cette section résume les travaux présentés dans [67].

Le but de cette section est d'établir formellement des liens entre plusieurs formalismes adaptés pour spécifier la dynamique qualitative des réseaux de régulation biologique, à savoir : le Graphe d'Interaction, référençant le signe de régulation directe entre les composants ; les Frappes de Processus, qui permettent de spécifier la dynamique qualitative du système avec divers degrés de connaissance des régulations conjointes ; et les paramètres de Thomas, qui spécifient complètement les fonctions qui régissent l'évolution des composants. L'autre motivation du travail présenté dans cette section est qu'il s'agit une étape importante pour aller plus loin dans l'étude des grands réseaux de régulation biologiques, et ce, au-delà de la simple analyse : il devient dès lors possible de contrôler en partie certains comportements. En effet, lorsque l'on veut modifier des comportements (par exemple, éviter l'accessibilité de certains états ou la création d'états stables), nous voudrions être en mesure d'obtenir en amont les résultats conformes au niveau réglementaire. Il en résulte la possibilité de faire des modifications sur le modèle de Frappes de Processus, puis d'obtenir le réseau de régulation biologique correspondant au modèle comportemental ainsi obtenu. Dans un cadre plus général, et c'est ce qui serait intéressant, nous serions capables de synthétiser les familles de réseaux de régulation biologique ayant certaines propriétés comportementales choisies.

D'abord, nous avons dérivé des règles qui, étant donné un modèle de Frappes de Processus, infèrent le Graphe d'Interaction correspondant à la dynamique du réseau de régulation biologique

codé. Le Graphe d'Interaction obtenu concerne uniquement les composants qui ont une influence réelle sur la dynamique. Il en résulte un Graphe d'Interaction contenant moins d'interactions que le Graphe d'Interaction de départ, puisque la spécification des évolutions des composants peut révéler des régulations non-fonctionnelles. À partir du Graphe d'Interaction dérivé, diverses analyses statiques permettent de conclure sur les propriétés globales de la dynamique du système (voir [143] pour un court résumé). Les plus connues sont les conjectures de Thomas (qui ont été prouvées depuis, par exemple dans [149, 152, 153] pour les réseaux booléens et discrets) qui se formulent comme suit :

- un circuit positif (c'est-à-dire comportant un nombre pair de régulations négatives) est une condition nécessaire à la présence de plusieurs points fixes ;
- un circuit négatif (c'est-à-dire comportant un nombre impair de régulations négatives) est une condition nécessaire à la présence d'oscillations soutenues dans la dynamique, et donc à la présence d'un attracteur constitué d'au moins deux états.

Ensuite, nous nous sommes attaqués à l'inférence des paramètres de Thomas compatibles avec un modèle de Frappes de Processus donné, c'est-à-dire qui précisent pleinement l'évolution des composants tout en respectant les coopérations spécifiées dans le modèle de Frappes de Processus. Plus formellement, la dynamique des réseaux de régulation biologique qui en résultent est assurée d'être incluse dans la dynamique du modèle de Frappes de Processus : toute transition dans le modèle (asynchrone) de Thomas existe dans le modèle de Frappes de Processus. En pratique, le nombre de paramètres de Thomas compatibles peut être énorme pour les grands réseaux où peu de coopérations sont spécifiées.

Cela met en évidence que les Frappes de Processus constituent un formalisme adapté pour analyser la dynamique de réseaux à grande échelle où la connaissance sur les coopérations est souvent incomplète : plutôt que de tenter de traiter un nombre éventuellement ingérable de modèles de Thomas, on peut déjà saisir les caractéristiques dynamiques précises à l'aide d'un modèle de Frappes de Processus unique rassemblant toutes les dynamiques plausibles.

Enfin, nous avons discuté d'une mise en œuvre de ces systèmes d'inférence utilisant la programmation par ensembles réponses (que nous désignerons ici par son appellation anglophone, à savoir Answer Set Programming (ASP) [12]). Nous avons choisi ASP pour les raisons suivantes :

- ASP traite efficacement la complexité inhérente aux modèles que nous étudions, nous permettant ainsi une exécution rapide des scripts nécessaires à nos algorithmes d'inférence ;
- ASP peut énumérer efficacement un large nombre d'ensembles réponses ;
- il est facile de contraindre les réponses en fonction de certaines propriétés.

Afin de donner une application détaillée de notre méthode, nous avons appliqué notre approche au modèle de contrôle de l'immunité du bactériophage lambda avec quatre composants et au modèle biologique du récepteur ERBB contrôlant le cycle cellulaire au niveau de la transition G1/S (contenant 20 composants). Cela a permis de montrer comment les actions conjointes peuvent



être définies à l'intérieur d'un modèle de Frappes de Processus pour affiner sa dynamique et d'étudier une classe plus précise de modèles de Thomas sous-jacents.

Notre travail est lié à l'approche de [91] qui s'appuie sur la logique temporelle, et à [45, 46] qui utilise la programmation par contraintes. Ces approches visent à déterminer une classe de modèles compatibles avec les données partielles disponibles sur la structure de régulation et les propriétés dynamiques. Notre méthode est fondée sur un modèle plutôt que sur des contraintes, ce qui permet de définir des propriétés sur la structure du système (telles que les coopérations entre composants). De plus, notre approche permet de traiter les larges réseaux biologiques. Enfin, il faut noter que, là, nous ne nous sommes pas intéressés à la dérivation des Frappes de Processus à partir d'un réseau de régulation biologique (précédemment décrite dans [139]) mais, au contraire, nous avons tâché de trouver un ensemble de réseaux de régulation biologique à partir de Frappes de Processus.

Les contributions présentées dans cette section s'appuient sur les résultats d'abord présentés dans [66], puis améliorés et étendus dans [67]. En plus de l'amélioration de l'efficacité et de la précision de l'inférence du Graphe d'Interaction, nous avons également considéré des régulations non-monotones, qui sont positives ou négatives selon le contexte. Cela permet d'appliquer notre méthodologie à une plus large classe de modèles. En outre, certaines parties ont été complètement réécrites afin d'expliquer plus précisément l'efficacité de l'implémentation d'ASP, et de donner une nouvelle application biologique détaillée de notre travail.

## **5.3 Reconstruction d'un réseau de régulation (réseau Booléen) et de ses règles logiques**

### **5.3.1 Les enjeux de l'apprentissage**

Grâce au développement récent de technologies de mesure à haut débit telles que les puces à ADN, les biologistes ont réussi à obtenir une grande quantité de profils d'expression génétique. Il devient alors crucial de pouvoir connecter les données et de construire un modèle prédictif du réseau de régulation génétique. L'analyse des réseaux biologiques exige souvent de s'appuyer sur un modèle approprié, mathématique ou calculatoire, pour représenter le système biologique. En raison de la complexité du système, les modèles supposent habituellement que la modification d'un nœud implique une activation immédiate (ou inhibition) de ses nœuds ciblés. Mais cette hypothèse est généralement incorrecte : certaines influences peuvent prendre du temps pour fonctionner, ce qui modifie le comportement du modèle. Ces influences avec retard peuvent jouer un rôle majeur dans divers systèmes biologiques, comme l'horloge circadienne chez les mammifères (tel qu'illustré par [43]) ou la réparation des lésions de l'ADN (tel que l'a montré [1]). Nous avons particulièrement besoin de capturer la mémoire du système, c'est-à-dire de garder trace des étapes précédentes.

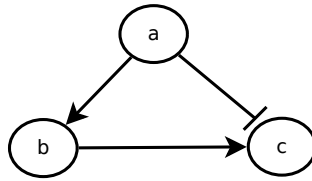


FIGURE 5.1 – Exemple d'une boucle de rétroaction

### 5.3.2 Modéliser les influences avec retard en réseaux booléens

Dans cette sous-section, nous introduisons les travaux présentés dans [150] et [151].

Les réseaux booléens se sont révélés de simples mais puissants formalismes pour modéliser et analyser la dynamique des réseaux de régulation génétique. La dynamique de réseaux booléens classiques est fondée sur l'hypothèse centrale que le délai de transmission des mises à jour des composants du réseau est homogène. Cela signifie que, pour des raisons de simplification, la modification d'un nœud résulte en une activation (ou inhibition) immédiate de ses nœuds cibles (comme étudié par exemple par [4]). Ceci ne paraît pas vraiment réaliste au sens où, dans un système réel biochimique, l'évolution se produit à différentes échelles de temps.

Le besoin d'intégrer des délais dans la modélisation est parfaitement illustré par le schéma de boucle de rétroaction. La boucle de rétroaction est un modèle de réseau qui apparaît dans de nombreux processus cycliques, par exemple *Escherichia coli* et *Yeast Saccharomyces cerevisiae* (tel que considéré par [96]). Dans des travaux tels que [118], des biologistes suggèrent que ces boucles jouent un rôle important dans l'amélioration du temps de réponse de la transcription des réseaux. Il se compose des éléments suivants (voir figure 5.1) : 3 gènes, notés  $a$ ,  $b$  et  $c$ , avec  $a$  régulant  $b$ ,  $b$  régulant  $c$ , et une régulation directe de  $a$  à  $c$ . Selon la nature des régulations (activations ou inhibitions ; la figure 5.1 considère arbitrairement une inhibition de  $a$  à  $c$ ) et leurs délais, la concurrence entre la régulation directe de  $a$  vers  $c$  et celle indirecte à travers  $b$  peut conduire à des comportements drastiquement différents. Pour analyser les boucles de rétroaction, les informations sur les délais respectifs des régulations en jeu sont cruciaux. Une modification, fût-elle faible, au niveau des délais peut conduire à des comportements complètement différents. L'analyse fine du comportement dynamique d'un tel modèle nécessite d'enrichir les modèles discrets classiques avec des délais.

Pour résoudre ce problème, différentes approches ont été conçues. On doit la plus connue à [163]. Pour comprendre précisément la dynamique de certains processus biologiques (comme le développement cellulaire) tout en prenant en compte la mémoire du système, les auteurs puisent leur inspiration dans les travaux sur les modèles de Markov. Ils ont introduit une extension des réseaux booléens, traditionnellement considérés comme un modèle Markov(1), à un modèle Markov( $k$ ), où  $k$  est le nombre de pas de temps pendant lequel un gène peut influencer un autre gène. Cette extension est appelée réseau booléen temporisé, abrégée en  $RBT(n, m, k)$ , avec  $n$  le nombre de gènes, l'expression de chaque gène au temps  $t + 1$  étant contrôlée par une

fonction booléenne des niveaux d'expression d'au plus  $m$  gènes aux temps  $\{t, t-1, \dots, t-(k-1)\}$ . Silvescu et al. ont même étendu le formalisme aux réseaux discrets multi-valués, appelés réseaux discrets temporels,  $RDT(n, m, k, D)$ , où un ensemble de niveaux  $0, 1, \dots, D-1$  est associé à chaque gène. Toutefois, leurs principaux résultats se concentrent sur les  $RBT(n, m, k)$  : ils proposent un algorithme d'apprentissage basé sur un arbre de décision qui infère un réseau booléen temporel à partir de données de séries temporelles. Ils ont aussi donné des limites sur la taille des données nécessaires pour déduire les réseaux booléens temporels. Ils ont illustré leurs résultats à travers des réseaux générés artificiellement, affirmant que la principale limite de leur méthode est le manque de données réelles sur de grands ensembles de données.

D'autres auteurs ont abordé l'idée de la modélisation avec retard et influences indirectes. Nous pouvons citer le travail de Chueh et Lu [39], qui ont étendu le formalisme de réseau booléen avec des retards. Pour modéliser le fait que l'induction d'une fonction booléenne peut ne pas activer immédiatement le gène ciblé, ils ont remplacé la relation déterministe classique entre la fonction booléenne et le gène ciblé par deux relations : (i) Une fonction de prérequis, représentant le fait que le statut activé d'un gène cible à  $t+1$  exige que la fonction booléenne à  $t$  soit vraie. (ii) Une fonction de similarité : la fonction booléenne et le gène cible sont dits similaires si l'état de la fonction booléenne et l'état du gène cible sont les mêmes. En d'autres termes, cela signifie que les opérateurs booléens classiques sont encodés dans la fonction de prérequis, tandis que la fonction de similarité permet de modéliser la précédence (sous forme de retards) entre les mises à jour simultanées.

Une autre approche pour modéliser les influences indirectes est donné dans le chapitre 6 de la thèse de Ghanbarnejad [74]. L'auteur rappelle que le temps qui s'écoule entre la production d'une molécule régulatrice et sa liaison à un site cible dépend à la fois de la molécule et de son site cible. C'est pourquoi il a décidé de considérer la dynamique de telle sorte que :

$$x_i(t) = f_i(x_1(t - \tau_{i1}), x_2(t - \tau_{i2}), \dots, x_N(t - \tau_{iN}))$$

avec :  $x_i$  la valeur du gène  $i$ ,  $t$  le point temporel considéré,  $\tau_{ij}$  le retard dans l'interaction entre un nœud source  $j$  et un nœud cible  $i$ . Pour ses recherches, l'auteur tire le retard  $\tau_{ij}$  sous forme d'un entier aléatoire obtenu à partir d'une distribution uniforme sur  $\{1, 2\bar{\tau} - 1\}$  pour chaque paire de nœuds  $i$  et  $j$ , le délai moyen  $\bar{\tau}$  étant un paramètre réglable. Cela permet d'introduire des réseaux booléens avec des retards distribués. La sémantique ici est synchrone, ce qui est très similaire à ce que nous visons dans cette section (mais avec la perspective d'étendre, ensuite, à la sémantique asynchrone).

Comme ces travaux dérivent du formalisme initial proposé par Silvescu et al. [163], nous considérons les  $RTB(n, m, k)$  et discutons d'un nouvel algorithme d'apprentissage.

### 5.3.3 Apprendre des réseaux booléens avec des influences retardées

Diverses approches ont été récemment conçues pour s'attaquer à la rétro-ingénierie des réseaux de régulation génétiques à partir des données d'expression. Cela a conduit à l'émergence de la biologie dite exécutable, dont le but est de fournir des méthodes formelles pour automatiquement synthétiser des modèles à partir d'expériences [97]. La plupart d'entre elles ne sont que statiques. Mais il y a un intérêt croissant pour les algorithmes d'inférence qui intègrent les aspects temporels. Koh et al. ont récemment étudié la pertinence de ces différents algorithmes [96]. Liu et al. ont proposé de déduire les réseaux de régulation génétiques avec retard via les réseaux bayésiens [102]. Lopes et Bontempi ont montré que les algorithmes d'inférence comprenant des caractéristiques temporelles sont plus performants que leurs équivalents statiques [103]. Le principal problème est alors d'être en mesure de déduire les délais temporels appropriés entre les influences en jeu. S'agissant d'un problème difficile, Zhang a affirmé que la question clé lors de l'analyse des données de séries temporelles consiste à segmenter les données de séries temporelles dans les différentes phases successives [173]. Sa contribution s'est concentrée ensuite sur la résolution de ce problème de segmentation et il a montré le bien-fondé de leurs approches sur différentes études de cas.

Toutes ces approches considèrent en entrée les données d'expression génétiques et en déduisent les régulations associées. Un problème usuel des approches discrètes prenant les données d'expression en entrée réside dans la détermination d'un seuil pertinent pour définir les états actif et inactif de l'expression d'un gène. Parmi le panorama d'approches permettant de traiter les données biologiques brutes, citons les travaux de certains auteurs, tels [164]. Dans cet article, Soinov et al. proposent une méthode originale qui ne considère pas un niveau de concentration, mais la façon dont la concentration est modifiée en présence/absence d'un régulateur. L'autre problème majeur en modélisation est la qualité des données d'expression. En d'autres termes, les données bruitées peuvent conduire à des erreurs dans le processus d'inférence. Par exemple, lorsqu'un gène est exprimé à un niveau faible, un faible rapport signal-bruit se traduirait par une mesure imprécise du comportement du gène.

Le pré-traitement des données est donc critique pour la pertinence des relations présumées entre les composants. Dans la suite de cette présentation, nous mettons de côté ce problème. Nous supposons que les données en entrée ont déjà été pré-traitées et ont abouti à un système fiable de transitions d'états.

En sus des problèmes intrinsèques de modélisation évoqués aux paragraphes précédents, les méthodes d'apprentissage existantes souffrent de limites en termes de temps et d'espace de calcul :

- En raison de la complexité du problème, la taille inférée du modèle est limitée : le réseau de régulation du gène inféré doit généralement être composé de moins de 15 composants et l'effet mémoire ne peut pas prendre en compte plus de  $k = 5$  étapes.
- De nombreuses approches échouent lorsque le réseau implique des interactions cycliques.

Temps d'exécution (# de secondes)						Taille de la sortie (# de règles)					
# Traces	k=1	k=2	k=3	k=4	k=5	# Traces	k=1	k=2	k=3	k=4	k=5
10	0,23	0,27	<b>0,16</b>	0,28	0,31	10	994	735	<b>413</b>	1.162	1.278
100	1,87	2,49	<b>1,63</b>	2,37	2,84	100	837	736	<b>383</b>	1.025	1.156
1.000	15,3	18,5	<b>13,3</b>	20,1	23,8	1.000	835	647	<b>364</b>	942	1.096
10.000	<b>146</b>	218	147	201	234	10.000	728	739	<b>408</b>	857	987
100.000	2.177	2.577	<b>1.643</b>	1.764	2.243	100.000	929	535	<b>390</b>	788	950
1.000.000	27.768	22.517	12.384	15.670	20.413	1.000.000	833	559	350	769	930

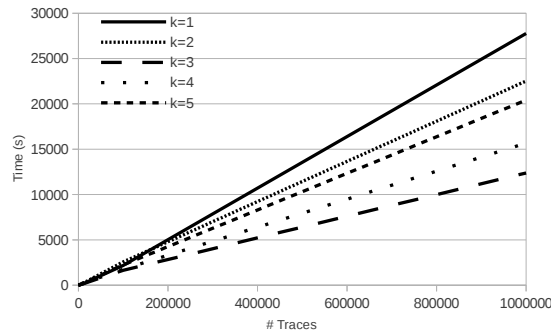


FIGURE 5.2 – Temps d'exécution de LFkT en fonction de la taille des données en entrée (en nombre de traces)

### 5.3.4 Données en entrée : traces d'exécution

Nous avons conçu une méthode logique d'apprentissage de tels modèles à partir de données de séries temporelles. Le détail de l'algorithme, baptisé **LFkT**, a été présenté dans [151]. Notre méthode prend en entrée un système de transitions d'états, qui dérive de la discrétisation préalable des données de séries temporelles. Pour le moment, nous ne nous préoccupons pas de la procédure de discrétisation et considérons qu'il s'agit d'une étape préliminaire effectuée en amont. Notre algorithme analyse alors les séquences d'états-transitions une par une, de manière itérative, pour construire un programme logique qui capture la dynamique de ces observations. Pour illustrer les mérites de cette approche, nous l'avons appliquée à l'apprentissage de données réelles issues de la littérature en bioinformatique. En utilisant les données du cycle cellulaire de la levure, nous allons synthétiser les résultats expérimentaux obtenus (résumés sur le tableau 5.2) et justifier le passage à l'échelle de notre méthode.

### 5.3.5 Apprentissage de réseaux booléens

### 5.3.6 Benchmarks

Nous avons appliqué cette approche à l'analyse de l'ensemble de données du cycle cellulaire de la levure initialement présentées dans [165] et [38], qui ont été par la suite analysées dans [101]. Dans ce dernier papier, Li et al. s'intéressent à l'inférence de réseaux de régulation génétiques à

### 5.3. RECONSTRUCTION D'UN RÉSEAU DE RÉGULATION (RÉSEAU BOOLÉEN) ET DE SES RÈGLES LOGIQUES

---

partir des données d'expression génétique temporelles. L'originalité de leur travail réside dans le fait qu'ils considèrent les retards dans les corrélations entre les gènes. Leur méthodologie est ainsi en mesure de prendre en compte des régulations entre gènes qui s'expriment avec retard. Les limites de l'approche sont que les auteurs ne considèrent que des paires d'influences les unes par rapport aux autres. Une autre limite de l'approche est qu'elle n'est pas en mesure de faire une distinction entre une régulation causale gène-gène et une coopération dans laquelle deux gènes, A et B, co-régulent un troisième gène C.

Dans nos expérimentations, nous avons considéré en entrée un ensemble de traces provenant du système du cycle cellulaire de la levure et avons effectué différentes expériences où nous avons changé, d'une part le nombre de traces en entrée, d'autre part la valeur de  $k$  (c'est-à-dire le nombre de pas temporels représentant la mémoire du système).

Le tableau 5.2 présente le temps d'exécution de l'algorithme **LFkT** pour apprendre les cinq différents réseaux booléens du cycle cellulaire de la levure proposés par [101]. Ces cinq modèles diffèrent par la "taille" de l'effet mémoire considéré (autrement dit le nombre  $k$  considéré). Ces cinq programmes se déclinent sous la forme de systèmes qui sont, respectivement, de Markov(1) à Markov(5). Dans ces expériences, pour chaque système, le nombre de variables est de seize et la longueur des traces en entrée est de cinq états. Toutes les expériences ont été effectuées avec une implémentation en C++ de **LFkT** sur un processeur Intel Xeon (X5650, 2.67GHz) avec 12Go de RAM. Le but principal de ces expériences est d'évaluer l'efficacité de notre approche, c'est-à-dire combien de traces l'algorithme **LFkT** peut gérer pour un  $k$  donné.

Dans le premier tableau de la figure 5.2, l'évolution du temps d'exécution de 10 à 1.000.000 traces (valeur supérieure arbitrairement choisie comme limite supérieure du passage à l'échelle de nos expériences) montre que, dans la pratique, l'apprentissage avec **LFkT** est linéaire dans le nombre de traces, lorsque le nombre de variables est fixe. Les résultats montrent que l'algorithme peut traiter plus d'un million de traces en moins de 10 heures. Lors de l'apprentissage sous forme de systèmes Markov(5), chaque trace peut être décomposée en 15 transitions interprétées car chaque trace est une séquence de cinq état-transition (une de 5 étapes, deux de 4 étapes, trois de 3 étapes, quatre de 2 étapes et cinq d'1 étape). Apprendre le programme Markov(5) à partir d'un million de traces d'exécution de taille cinq exige le traitement de 15 millions de transitions interprétées. Apprendre les programmes Markov(4) à Markov(1) requiert de traiter respectivement 14 millions, 12 millions, 9 millions et 5 millions de transitions interprétées. Intuitivement on peut s'attendre à ce que l'apprentissage de système Markov(2) prenne beaucoup plus de temps que l'apprentissage de système Markov(1). Mais chaque programme est différent, c'est-à-dire que le programme Markov(2) n'est pas une extension du programme de Markov(1) avec des règles de 2 étapes. C'est pourquoi l'exécution n'est pas toujours plus coûteuse pour un plus grand  $k$  : le temps d'apprentissage dépend aussi des règles qui sont apprises. Dans ce jeu d'expériences, le meilleur temps d'exécution est obtenu avec le programme Markov(3). Nous ne pouvons pas dire que les règles de ce programme sont plus simples que les autres, mais elles sont

plus simples à apprendre pour l'algorithme. Dans le deuxième tableau, nous constatons que le nombre de règles apprises pour le programme Markov(3) est nettement plus faible que pour les autres. Cela signifie que l'algorithme a besoin de moins de règles pour comparer chaque trace, ce qui peut expliquer la plus grande vitesse.

Dans cette étude, afin d'être fidèles aux expériences biologiques présentées par [101], nous avons considéré  $k = 5$  comme maximum. Mais notre algorithme réussit à traiter de plus grands effets mémoire. Sur certains exemples factices aléatoires, nous avons pu apprendre des systèmes Markov(7) avec les performances suivantes : nous pouvons apprendre 10 traces en 2.8s, 100 traces dans 27s, 1000 traces en 249s, 10.000 traces de 3,621s, 100 000 traces en 39,973s, 1.000.000 traces dans 441,270s. Même si le temps de calcul augmente, il faut garder à l'esprit que notre méthode est conçue pour permettre les raffinements successifs de l'effet mémoire d'un modèle. Ces résultats montrent que cette approche est traitable même avec un grand nombre de traces en entrée.

## 5.4 Promesses de l'apprentissage automatique pour la biologie

Nos travaux actuels et futurs visent à adapter l'approche développée dans ce chapitre directement aux données produites par les biologistes. Cela nécessite de se connecter à différentes bases de données biologiques afin d'en extraire les données de la série temporelles et de les exploiter pour apprendre les familles de réseaux de régulation génétiques sous-jacents (sous une forme pouvant être, comme dans la section précédente, celle d'un programme logique). En raison du bruit inhérent aux données biologiques, il est fondamental de trouver comment effectuer une discrétisation efficace des données ou d'inclure la notion de bruit à l'intérieur du cadre de la modélisation. Nous comptons donc discuter de la procédure de discrétisation et de la robustesse de notre approche par rapport aux données bruitées et la comparer aux approches existantes, par exemple bayésiennes [14].

En ce qui concerne le formalisme de modélisation, nous envisageons d'étendre la méthodologie à une sémantique asynchrone, comme celle des Frappes de Processus.

Enfin, nous aborderons également les réseaux multi-valués qui peuvent être utiles pour capturer des comportements ne pouvant être représentés dans un cadre booléen.

## 5.5 Synthèse

### 5.5.1 Inférence des relations logiques entre gènes

Les travaux présentés dans ce chapitre nous ont d'abord permis d'établir la relation abstraite entre les Frappes de Processus et les approches de Thomas pour une modélisation qualitative des réseaux de régulation biologique. Les Frappes de Processus permettent une représentation abstraite de la dynamique d'un réseau de régulation biologique (tout en permettant de gérer

le cas où nous n'avons qu'une connaissance incomplète des coopérations entre les composants) qui ne peut pas être exactement représentée dans le formalisme de René Thomas par une seule instance de paramétrage de réseau de régulation biologique. Cela motive la concrétisation de modèles de Frappes de Processus en un ensemble de modèles de Thomas compatibles afin de bénéficier des avantages complémentaires de ces deux formalismes.

Nous avons d'abord proposé une méthode originale d'inférence du Graphe d'Interaction à partir d'un réseau de régulation biologique lorsque sa dynamique est spécifiée dans le cadre du paradigme des Frappes de Processus. Un Graphe d'Interaction donne une représentation abstraite compacte de l'influence des composants entre eux. Ensuite, en nous basant sur une inférence préalable de la paramétrisation du modèle de René Thomas pour les réseaux de régulation biologique via un modèle de Frappes de Processus, nous avons délimité l'ensemble des paramétrisations de Thomas admissibles qui sont compatibles avec la dynamique sous forme de Frappes de Processus, et nous avons donné des arguments concernant leur correction [67]. Une paramétrisation est compatible avec les Frappes de Processus si ses dynamiques (en termes de transitions possibles) sont incluses dans la dynamique de Frappes de Processus. L'énumération de telles paramétrisations a pu être abordée de manière efficace en utilisant Answer Set Programming. Nous avons illustré la méthode globale avec plusieurs résultats sur de petits et de grands modèles biologiques.

Nous envisageons plusieurs extensions aux travaux présentés ici. D'abord, le lien entre les raffinements successifs menés au cours du processus de modélisation mérite d'être étudié plus formellement. En effet, il est prometteur d'affiner un modèle de Frappes de Processus en enlevant ou en ajoutant des actions : l'étude et la formalisation d'un tel processus permettrait de prédire les changements de comportement et conduire à des modèles plus précis. C'est l'objet de travaux que nous avons entamés via la thèse d'Emna Ben Abdallah, et la collaboration avec Tony Ribeiro et Katsumi Inoue. Ensuite, l'inférence des réseaux de régulation biologique exprimés sous la forme de multiplexes [20] peut être intéressante en pratique car ce modèle permet de réduire implicitement les paramétrages possibles en faisant apparaître les coopérations au niveau du Graphe d'Interaction. En raison de leur atomicité, les Frappes de Processus permettent de spécifier une gamme de coopérations qui ne peut être complètement capturée par une seule instance de multiplexes. Dès lors, nous pourrions mener l'inférence d'un ensemble de multiplexes compatibles. Enfin, afin d'améliorer les performances dans l'inférence du Graphe d'Interaction, nous envisageons d'explorer des opérations de projection sur la structure des Frappes de Processus pour défaire les coopérations entre les composantes et réduire la cardinalité des configurations à explorer en rendant les interactions indépendantes.

### 5.5.2 Apprendre des influences retardées au sein de systèmes biologiques

Nous nous sommes concentrés sur la conception d'une approche logique pour apprendre un réseau de régulation génétique avec influences retardées à partir d'une connaissance pré-existante



s'exprimant sous la forme d'un système de transition d'états. Tel que mentionné précédemment, notre approche suppose qu'il y a eu un pré-traitement des données de séries temporelles.

Alors que les travaux existants menés dans le domaine des approches logiques ne permettaient pas de saisir les influences retardées entre les composants biologiques [85, 88], notre approche construit un programme logique qui capture la dynamique avec retard d'un système Markov( $k$ ).

Nous basant sur le cadre proposé par [87] pour apprendre des programmes logiques à partir de système de transition d'états, nous avons étendu ces travaux : alors que l'algorithme précédent ne portait que sur des transitions en une étape (c'est-à-dire que l'approche supposait que l'état du système à l'instant  $t$  dépend uniquement de son état à l'instant  $t - 1$ ), nous avons proposé une approche qui examine  $k$  successions d'états-transitions (séquence d'au plus  $k$  transitions d'états). Nous sommes ainsi désormais en mesure de saisir l'influence avec retard dans une méthodologie de *programmation logique inductive*.

Nous travaillons actuellement sur les extensions de ces travaux avec d'autres sémantiques de mises à jour de l'état des composants. C'est ainsi que nous avons récemment considéré une extension aux modèles multi-valués d'une part, asynchrones non-déterministes d'autre part. Ces travaux font actuellement l'objet de soumissions d'articles.

## CONCLUSION GÉNÉRALE

## 6.1 Contribution

Arrivés à cette étape du manuscrit, il serait tentant d'écrire que les travaux présentés dans ce document ont émergé du croisement fertile entre des approches formelles et des champs d'application issus de la biologie. Mais est-il correct de considérer la biologie seulement comme un terrain d'application ? Elle est en fait plus que cela. De la même manière que l'informatique ne saurait se réduire à un outil permettant d'automatiser des tâches nécessaires au biologiste, il est vain de ne considérer la biologie qu'au travers du prisme des cas d'étude qu'elle peut fournir au chercheur en informatique.

C'est véritablement le dialogue entre ces deux champs disciplinaires qui est non seulement stimulant intellectuellement, mais aussi apte à faire émerger de nouvelles approches fructueuses pour les deux domaines. Nous avons essayé de mener, modestement, nos recherches avec cette idée en tête. L'introduction même des Frappes de Processus préside à ce questionnement. Plutôt que de prolonger l'étude de modèles existants, nous avons préféré nous demander comment modéliser, de manière atomique, les interactions entre composants biologiques. Pour cela, nous sommes repartis de considérations basiques sur ces interactions tout en investiguant l'apport de formalismes tels que le  $\pi$ -calcul pour capturer ces interactions [121, 140].

Cette analyse nous a permis de faire émerger un nouveau formalisme [139], les Frappes de Processus, sur lesquels nous avons été capables de développer des méthodes efficaces d'analyse. Parmi ces méthodes figure l'analyse des points fixes ou de propriétés d'accessibilités successives et ce, même dans des réseaux de grande taille [142]. Les Frappes de Processus reposent sur l'idée de représenter de manière atomique les évolutions possibles d'un processus (qui représente un composant donné à un certain niveau) activé par la frappe d'au plus un autre processus dans le

système. Le travail de modélisation avec les Frappes de Processus repose sur l'idée de commencer par construire la dynamique la plus large possible, pour la restreindre progressivement à l'aide des données biologiques à disposition. L'ajout de données de coopérations sur le modèle, représentant par exemple la manière dont l'action conjointe de deux composants s'opère sur un troisième, en est un exemple. Pour affiner les modèles, mais aussi dans l'idée d'ajouter des données temporelles quantitatives dans le formalisme, nous avons ensuite enrichi ce formalisme afin d'y intégrer des classes de priorité [64, 68].

Même si les Frappes de Processus pourraient être vues comme "un modèle de plus", nous avons pris soin - à toutes les étapes de leur enrichissement - de les positionner par rapport aux modèles existants. Ainsi, tout autant que nous proposons une nouvelle approche pour l'inférence de paramètres discrets dans les réseaux de Thomas, nous avons établi un lien formel, dans les deux sens, entre les Frappes de Processus et ce modèle [66]. De même, nous avons mis en évidence les liens entre les Frappes de Processus et les réseaux de Petri [141], relations étendues ensuite aux Frappes de Processus avec classes de priorités et réseaux de Petri à arcs inhibiteurs [64].

En parallèle, nous avons cherché à étendre le spectre scientifique de nos approches. Si nos premiers efforts ont porté sur le développement, pour la biologie, de méthodes formelles inspirées du model-checking, nous avons élargi ce cadre à des raisonnements issus de l'apprentissage automatique. Nous avons d'abord exploré l'apport de la programmation logique, et plus spécifiquement de l'Answer Set Programming (ASP), dans l'inférence des paramètres de réseaux de régulation [66]. Nous nous sommes ensuite penchés sur les algorithmes d'inférence pouvant, à partir de données d'expression chronologiques, permettre la reconstruction de modèles. Ces travaux constituent les briques indispensables pour les développements que nous comptons désormais mener, et que nous présentons dans la section suivante.

## 6.2 Perspectives

Avec l'essor concomitant des sciences et technologies de l'information et de la communication et de nouvelles techniques de mesures, il est désormais possible d'accéder à un large volume de données biologiques [119].

Différentes approches ont d'ores et déjà vu le jour pour traiter la rétro-ingénierie des réseaux de régulation à partir de profils d'expression génétiques [77, 101, 163]. Mais la plupart des méthodologies existantes ne traitent que des modèles statiques. Cela veut ainsi dire, par exemple, que toute l'information contenue dans des données de séries temporelles n'est pas encore exploitée. Dans ce contexte, nous souhaitons notamment traiter toute la richesse et la qualité des données biologiques et donc nous attaquer notamment à la richesse des données de séries temporelles. Les informations temporelles quantitatives qu'elles contiennent peuvent en effet être utiles pour comprendre la dynamique de nombreux systèmes.

Le traitement des données de séries temporelles pose plusieurs défis auxquels nous comptons

nous attaquer en combinant des approches logiques et de vérification formelle.

### **6.2.1 Apprentissage de réseaux de régulation biologiques modélisant des comportements complexes**

D'abord, l'inférence automatique de modèles dynamiques à partir des données biologiques est un enjeu majeur pour étudier finement des systèmes où le temps joue un rôle critique, telle l'horloge circadienne des mammifères [43] ou le mécanisme de réparation de l'ADN par p53 [1].

Comme nous l'avons détaillé dans le corps du manuscrit, la collaboration avec le National Institute of Informatics de Tôkyô nous a ouvert de nouvelles pistes dans le domaine des approches issues de la logique et de l'apprentissage automatique. Cette perspective est prometteuse, notamment dans notre souhait de ne plus uniquement construire les modèles de Frappes de Processus "à la main", mais de pouvoir directement nous brancher à des bases de données biologiques pour en produire et valider les modèles associés.

Nous nous sommes récemment intéressés à l'information chronologique des données de séries temporelles. Ainsi, dans [151], nous avons introduit une approche logique pour apprendre des réseaux booléens avec des retards à partir de données de séries temporelles. Toutefois, cette approche ne tient pas encore compte des informations temporelles quantitatives entre les changements d'états du modèle. Nous avons commencé à travailler sur une extension de cette approche à d'autres contextes de modélisation (modèles multi-valués, sémantique asynchrone) et nous visons maintenant à étendre cette approche afin de préserver l'expressivité des données de séries temporelles en termes de délais. Dès lors, nous pourrions exporter les résultats sous la forme d'un modèle étendu (avec des délais) de Frappes de Processus et développer de nouvelles méthodes d'analyse à grande échelle. Il sera notamment intéressant d'étudier comment les méthodes mises au point dans le contexte des Frappes de Processus discrètes, ou enrichies avec des priorités, pourront être étendues dans le cas de modèles temporisés.

### **6.2.2 Révision de modèles**

Des expériences additionnelles peuvent apporter des précisions sur un modèle (rendant ainsi nécessaire un raffinement de celui-ci) ou permettre de discriminer les modèles acceptables parmi un ensemble de modèles considérés, pourvu que le bruit associé ne remette pas en cause la précision et la pertinence de l'information correspondante.

La production intensive de nouvelles données biologiques, ainsi que le nombre croissant de données de séries temporelles, pose un autre défi : comment réviser efficacement un modèle existant (construit par exemple à partir de données précédemment obtenus et/ou de l'expertise de collaborateurs biologistes) avec de nouvelles données ?

Cela rend impératif la conception de méthodes de révision efficaces, capables de mettre à jour la connaissance préalable sur le système avec des informations additionnelles sur la dynamique. Autrement dit, il y a un besoin fort pour des méthodes automatiques de mise à jour d'un modèle

de manière à ce que celui-ci soit cohérent avec les observations et un ensemble de critères donnés (par exemple, minimiser le nombre de modifications sur le modèle).

La révision et la complétion (le premier s'entend par l'ajout ou la suppression d'éléments dans le modèle, tandis que le second exclut toute suppression) ont fait l'objet de nombreux travaux récents. Dans [6], les auteurs se sont attaqués à la complétion de réseaux booléens stationnaires. Ils ont raffiné leur méthode au fil des années. Leurs derniers travaux [125] se concentrent sur la complétion dans les *Time Varying Genetic Networks*. Dans ces réseaux, la topologie reste constante au fil du temps, mais la nature des interactions entre les composants (activation, inhibition, ou aucune interaction) peut changer en un nombre (fini) de points. Ces approches de complétion (qui, spécifiquement chez ces auteurs, englobe à la fois la suppression et l'addition d'interactions) ont été appliquées avec succès à des cas d'étude biologiques, par exemple les données du challenge DREAM4 [127]. De plus, l'implantation logicielle a été améliorée grâce à des heuristiques [126]. Ces méthodes souffrent toutefois de certaines limites : elles bornent généralement le nombre d'interactions entrantes sur un composant biologique (afin de limiter l'explosion combinatoire) et se restreignent aux réseaux acycliques.

Des travaux dans le domaine de la logique ont été également menés pour traiter le problème de la révision des réseaux. C'est ainsi que des approches logiques ont été mises au point dans le cas de réseaux de causalité [86] et de réseaux moléculaires représentés avec le langage SBGN-AF [172].

Même si elle constitue un champ de recherche à part entière, la révision est évidemment fortement liée à l'inférence de modèles. Partant d'un modèle vide, il est en effet possible d'utiliser la révision pour construire incrémentalement un modèle pertinent. En particulier, Answer Set Programming - qui a, on l'a vu, fait ses preuves dans différents contextes de représentation des connaissances et de raisonnement automatique [12, 13, 129] - s'est avéré utile pour la reconstruction de réseaux [56] et l'inférence de réseaux métaboliques [170].

À notre connaissance, il n'y a pas eu, à l'heure actuelle, de travaux portant sur la révision de modèles temporisés sans faire d'hypothèse restreignant la structure du réseau.

Dans ce contexte, notre objectif est de construire des approches logiques efficaces permettant la révision de modèles qualitatifs de systèmes biologiques dynamiques tels que les réseaux de régulation génétiques. Différents cas de figure pourront être envisagés. Dans un premier temps, nous supposons que l'ensemble des composants en interaction est fixé, et considérerons les ajouts/suppressions potentiels d'interactions entre les composants. Dans un second temps, nous pourrions étudier l'impact que représente l'ajout d'un composant (correspondant par exemple à un gène dont on pensait initialement, à tort, qu'il n'avait pas d'influence), voire la suppression. L'originalité de ces travaux sera de traiter ce problème dans un contexte hybride, avec des délais quantitatifs représentant le temps entre l'activation d'une interaction et le moment où son effet est observable.

Nous prévoyons également de traiter les données qui seraient incohérentes avec des connais-

sances précédentes. C'est-à-dire qu'étant donné un modèle existant et une (ou plusieurs) nouvelle(s) observation(s), nous souhaitons construire automatiquement l'ensemble minimal d'actions manquantes qui peuvent capturer l'observation tout en préservant la dynamique du modèle.

Enfin, quand plusieurs modèles sont compatibles avec des données de séries temporelles, nous souhaitons proposer des algorithmes semi-automatiques permettant de décider, d'un point de vue du modèle, quel type d'expériences pourraient être accomplies pour discriminer efficacement les modèles. Cette question est cruciale car, d'un point de vue pratique, les expériences biologiques sont coûteuses et/ou ne permettent pas d'observer toutes les variables du système. La notion d'observabilité, importante au sein de la communauté de l'automatique, est là aussi primordiale.

Enfin, nous devons aussi nous pencher sur le bruit intrinsèque aux données biologiques, notamment pour caractériser la robustesse d'un modèle. Ce travail est d'ailleurs lié au point suivant.

### 6.2.3 Étude de dynamiques complexes

Afin de discriminer les modèles pertinents parmi une famille de modèles existants, il est nécessaire de poursuivre la conception de méthodes efficaces d'analyse des propriétés clefs de ces modèles.

Afin d'affiner la qualité d'un modèle, il est crucial d'établir un pont entre les propriétés observées du système biologique et la dynamique du modèle. Cela signifie que nous devons mener une procédure d'élicitation du modèle vis-à-vis d'un certain nombre de propriétés, en utilisant par exemple une démarche basée sur le model-checking. Dans ce cas, les propriétés doivent être exprimées avec une logique modale, par exemple des extensions de LTL [147] et CTL [41] à l'instar de TCTL, utile pour la vérification de propriétés temporelles quantitatives [7].

Lorsque nous analysons des systèmes dynamiques, un large spectre de propriétés sont intéressantes à étudier : états stables, accessibilité, oscillations, etc. Récemment, nous nous sommes intéressés aux propriétés associées au concept de résilience [76, 82]. Intuitivement, ce terme représente la capacité d'un système à absorber des changements soudains de l'environnement (social, écologique, médical, etc.) tout en maintenant ses fonctionnalités clefs. Ce concept est intéressant dans de nombreux secteurs : dans le domaine sociétal (pour garantir, par exemple, le bon fonctionnement d'une ville ou d'un pays après un séisme dévastateur), en psychologie (quand il s'agit de se remettre d'un traumatisme), en biologie, etc.

La notion de résilience manque, la plupart du temps, d'une définition formelle. Certains auteurs la considèrent uniquement sous l'angle de la stabilisabilité ou de la robustesse, quand d'autres la distinguent de ces notions. Depuis quatre ans (et notamment depuis le tremblement de terre du 11 mars 2011 au Japon), la "résilience des systèmes" est un axe important de recherche transdisciplinaire. Un des objectifs de nos travaux récents a été de définir un ensemble de principes unifiés pour concevoir des systèmes résilients [159] à travers la définition d'un cadre général et flexible, applicable à un large spectre de spécifications. Nous avons conçu un formalisme

qui constitue un pont entre la programmation par contrainte et les modèles dynamiques. Nous avons alors défini formellement la notion de résilience [159] en nous basant sur la littérature. Cela nous a conduit à compiler ce principe de résilience sous la forme de différents facteurs, principalement la résistance, la recouvrabilité et la fonctionnalité du système.

Notre travail consiste désormais à dresser les ponts entre ces propriétés et la logique modale, de manière à pouvoir automatiquement vérifier les propriétés de résilience. L'étape suivante consistera à s'inspirer de la théorie du contrôle pour suggérer des modifications sur le système afin de garantir que ces propriétés seront satisfaites. Par ailleurs, à l'heure actuelle, ces travaux se limitent pour le moment à des modèles chronologiques. Nous avons récemment proposé une extension hybride à cette méthodologie, que nous comptons développer de manière plus approfondie. En effet, comme nous l'avons évoqué précédemment, le temps quantitatif sera nécessaire pour représenter certaines caractéristiques, par exemple la résilience de l'horloge circadienne des mammifères (permettant d'"absorber" les décalages horaires lors de vols long-courrier). D'autres propriétés dynamiques nécessitent également cette notion de temps quantitatif, telle la notion de fenêtre de vulnérabilité (qui a été abordée dans des études sur la sécurité informatique [24]).

#### **6.2.4 Champs d'application : vers la médecine personnalisée**

Si nous tentons de nous projeter à plus long terme, la médecine personnalisée est un des enjeux majeurs des recherches en bio-informatique. Demain, avec l'apparition de séquenceurs ADN portables (tel le séquenceur USB MinION) et autres dispositifs de *quantified self*, nous disposerons potentiellement d'une phénoménale quantité de données sur chaque individu. D'une exploitation efficace de celles-ci dépendra la conception de traitements médicaux novateurs, adaptés à chacun. Il est en effet possible d'imaginer qu'un individu malade pourra, dans certains cas, être soigné par des médicaments (voire des nano-machines ou des cellules vivantes) spécialement créés (par exemple via des mécanismes d'impression 3D, tel ce projet de l'Université de Tôkyô et de Fuji Film qui travaillent sur une imprimante 3D générant de la peau artificielle) après analyse de son génome. Évidemment, les problèmes sanitaires, tout comme les questions légales et éthiques derrière une telle démarche, ne sont pas minces. Les questions scientifiques non plus. Il sera dès lors crucial d'être en mesure d'analyser les mécanismes biologiques non plus en considérant les différentes échelles - biochimiques, génétiques, cellulaires - indépendamment les unes des autres, mais de considérer l'ensemble de l'organisme dans sa complexité systémique. Le développement d'approches adaptées pour traiter des systèmes multi-échelles (ou des systèmes couplés) est primordial.

Sur un horizon plus court, nous comptons illustrer nos perspectives dans le domaine de l'inférence et de la révision à des données réelles, à commencer par l'horloge circadienne des mammifères, qui fait l'objet d'un projet ANR dans lequel nous sommes impliqués et qui se place comme le point de départ de travaux plus approfondis sur le sujet. Partant d'un modèle

dynamique qui a été constitué lors d'un projet exploratoire PEPH, les méthodes d'inférence et de révision trouveront là un terreau fertile de mise en œuvre.





## BIBLIOGRAPHIE

- [1] W. ABOU-JAOUDÉ, D. A. OUATTARA, AND M. KAUFMAN, *From structure to dynamics : frequency tuning in the p53–mdm2 network : I. logical approach*, Journal of theoretical biology, 258 (2009), pp. 561–577.
- [2] J. AHMAD, G. BERNOT, J.-P. COMET, D. LIME, AND O. ROUX, *Hybrid modelling and dynamical analysis of gene regulatory networks with delays*, ComPlexUs, 3 (2006), pp. 231–251.
- [3] T. AKUTSU, M. HAYASHIDA, AND T. TAMURA, *Algorithms for inference, analysis and control of Boolean networks*, Springer, 2008.
- [4] T. AKUTSU, S. KUHARA, O. MARUYAMA, AND S. MIYANO, *Identification of genetic networks by strategic gene disruptions and gene overexpressions under a boolean model*, Theoretical Computer Science, 298 (2003), pp. 235–251.
- [5] T. AKUTSU, A. A. MELKMAN, T. TAMURA, AND M. YAMAMOTO, *Determining a singleton attractor of a boolean network with nested canalizing functions*, Journal of Computational Biology, 18 (2011), pp. 1275–1290.
- [6] T. AKUTSU, T. TAMURA, AND K. HORIMOTO, *Completing networks using observed data*, in Algorithmic Learning Theory, Springer, 2009, pp. 126–140.
- [7] R. ALUR, C. COURCOUBETIS, AND D. DILL, *Model-checking for real-time systems*, in Logic in Computer Science, 1990. LICS'90, Proceedings., Fifth Annual IEEE Symposium on Logic in Computer Science, IEEE, 1990, pp. 414–425.
- [8] A. ANDREYCHENKO, M. MAGNIN, AND K. INOUE, *Modeling of resilience properties in oscillatory biological systems using parametric time petri nets*, in 13th International Conference on Computational Methods in Systems Biology - , CMSB 2015, Nantes, France, O. F. Roux and J. Bourdon, eds., vol. 9308 of Lecture Notes in Computer Science, Springer, 2015.
- [9] J. ARACENA, E. G. CH., A. MOREIRA, AND L. SALINAS, *On the robustness of update schedules in boolean networks.*, Biosystems, 97 (2009), pp. 1–8.

- [10] P. BALDAN, N. BUSI, A. CORRADINI, AND G. MICHELE PINNA, *Functional concurrent semantics for petri nets with read and inhibitor arcs*, in CONCUR 2000 — Concurrency Theory, C. Palamidessi, ed., vol. 1877 of Lecture Notes in Computer Science, Springer Berlin Heidelberg, 2000, pp. 442–457.
- [11] A.-L. BARABASI AND Z. N. OLTVAI, *Network biology : understanding the cell's functional organization*, Nature Reviews Genetics, 5 (2004), pp. 101–113.
- [12] C. BARAL, *Knowledge Representation, Reasoning and Declarative Problem Solving*, Cambridge University Press, 2003.
- [13] C. BARAL, *Using answer set programming for knowledge representation and reasoning : Future directions*, in ICLP, 2008, pp. 69–70.
- [14] N. A. BARKER, C. J. MYERS, AND H. KUWAHARA, *Learning genetic regulatory network connectivity from time series data*, Computational Biology and Bioinformatics, IEEE/ACM Transactions on, 8 (2011), pp. 152–165.
- [15] G. BATT, *Design, optimization and control in systems and synthetic biology*, habilitation à diriger des recherches, Université Paris-Diderot - Paris VII, Mar. 2014.
- [16] G. BATT, M. PAGE, I. CANTONE, G. GOESSLER, P. MONTEIRO, AND H. DE JONG, *Efficient parameter search for qualitative models of regulatory networks using symbolic model checking*, Bioinformatics, 26 (2010), pp. i603–i610.
- [17] E. BEN ABDALLAH, M. FOLSCHETTE, O. ROUX, AND M. MAGNIN, *Exhaustive analysis of dynamical properties of biological regulatory networks with answer set programming*, in 2015 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), IEEE, 2015, pp. 281–285.
- [18] D. BÉRENGUIER, C. CHAOUÏYA, P. MONTEIRO, A. NALDI, E. REMY, D. THIEFFRY, AND L. TICHIT, *Dynamical modeling and analysis of large cellular regulatory networks*, Chaos : An Interdisciplinary Journal of Nonlinear Science, 23 (2013), p. 025114.
- [19] L. BERNARDINELLO AND F. DE CINDIO, *A survey of basic net models and modular net classes*, in Advances in Petri Nets 1992, G. Rozenberg, ed., vol. 609 of Lecture Notes in Computer Science, Springer Berlin / Heidelberg, 1992, pp. 304–351.
- [20] G. BERNOT, J.-P. COMET, AND Z. KHALIS, *Gene regulatory networks with multiplexes*, in European Simulation and Modelling Conference Proceedings, 2008, pp. 423–432.
- [21] G. BERNOT, J.-P. COMET, A. RICHARD, AND J. GUESPIN, *Application of formal methods to biological regulatory networks : extending thomas bi asynchronous logical approach with temporal logic*, Journal of theoretical biology, 229 (2004), pp. 339–347.

- 
- [22] N. BONZANNI, K. A. FEENSTRA, W. FOKKINK, AND J. HERINGA, *Petri nets are a biologist's best friend*, in Formal Methods in Macro-Biology, Springer, 2014, pp. 102–116.
  - [23] D. BRAND AND P. ZAFIROPULO, *On communicating finite-state machines*, Journal of the ACM, 30 (1983), pp. 323–342.
  - [24] E. BURSZTEIN AND J. GOUBAULT-LARRECQ, *A logical framework for evaluating network resilience against faults and attacks*, in Advances in Computer Science–ASIAN 2007. Computer and Network Security, Springer, 2007, pp. 212–227.
  - [25] S. CAROLAN, F. CHINESTA, C. EVAIN, M. MAGNIN, AND G. MOREAU, *Towards augmented learning in science and engineering in higher education*, in IEEE 13th International Conference on Advanced Learning Technologies (ICALT), IEEE, 2013, pp. 512–513.
  - [26] S. CAROLAN, F. CHINESTA, C. EVAIN, M. MAGNIN, AND G. MOREAU, *Extracting domain ontologies from reference books*, in IEEE 14th International Conference on Advanced Learning Technologies (ICALT), IEEE, 2014, pp. 544–545.
  - [27] S. CAROLAN AND M. MAGNIN, *Le mooc et la motivation : les élèves face à une formation autogérée*, in 6ème Conférence Environnements Informatiques pour l'Apprentissage Humain (EIAH 2013), Atelier MOOC, 2013.
  - [28] S. CAROLAN, M. MAGNIN, AND J.-M. GILLIOT, *Engineering moocs for future engineers : Integrating moocs into formal learning environments*, EMOOCs 2014 : 2nd MOOC European stakeholders summit, (2014).
  - [29] S. CAROLAN, M. MAGNIN, J.-M. GILLIOT, A.-C. GROLLEAU, AND C. VAUFREY, *Un mooc et ses partenaires*, in 9ème Colloque International Technologies de l'Information et de la Communication pour l'Enseignement (TICE 2014). Retour d'Expérience (ReX), 2014.
  - [30] S. CAROLAN, M. MAGNIN, AND A.-L. KABALU, *Sparkling a digital revolution : Digital educational tools in fragile and emerging learning contexts*, in Digital Intelligence 2014, 2014.
  - [31] C. CHANCELLOR, A. AMMAR, F. CHINESTA, M. MAGNIN, AND O. ROUX, *Linking discrete and stochastic models : The chemical master equation as a bridge between process hitting and proper generalized decomposition*, in Computational Methods in Systems Biology, A. Gupta and T. Henzinger, eds., vol. 8130 of Lecture Notes in Computer Science, Springer, 2013, pp. 50–63.
  - [32] C. CHAOUIYA, H. KLAUDEL, AND F. POMMEREAU, *A modular, qualitative modeling of regulatory networks using petri nets*, in Modeling in Systems Biology, Springer, 2011, pp. 253–279.

- [33] C. CHAOUIYA, A. NALDI, E. REMY, AND D. THIEFFRY, *Petri net representation of multi-valued logical regulatory graphs*, Natural Computing, 10 (2011), pp. 727–750.
- [34] C. CHAOUIYA, E. REMY, B. MOSSÉ, AND D. THIEFFRY, *Qualitative analysis of regulatory graphs : a computational tool based on a discrete formal framework*, in Positive Systems, Springer, 2003, pp. 119–126.
- [35] C. CHAOUIYA, E. REMY, AND D. THIEFFRY, *Petri net modelling of biological regulatory networks*, J. of Discrete Algorithms, 6 (2008), pp. 165–177.
- [36] C. CHETTAOUI, F. DELAPLACE, P. LESCANNE, M. VESTERGAARD, AND R. VESTERGAARD, *Rewriting game theory as a foundation for state-based models of gene regulation*, in Computational Methods in Systems Biology, International Conference, CMSB 2006, Trento, Italy, October 18-19, 2006, Proceedings, 2006, pp. 257–270.
- [37] F. CHINESTA, M. MAGNIN, O. ROUX, A. AMMAR, AND E. CUETO, *Kinetic theory modeling and efficient numerical simulation of gene regulatory networks based on qualitative descriptions*, Entropy, 17 (2015), pp. 1896–1915.
- [38] R. J. CHO, M. J. CAMPBELL, E. A. WINZELER, L. STEINMETZ, A. CONWAY, L. WODICKA, T. G. WOLFSBERG, A. E. GABRIELIAN, D. LANDSMAN, D. J. LOCKHART, ET AL., *A genome-wide transcriptional analysis of the mitotic cell cycle*, Molecular cell, 2 (1998), pp. 65–73.
- [39] T.-H. CHUEH AND H. H.-S. LU, *Inference of biological pathway from gene expression profiles by time delay boolean networks*, PloS one, 7 (2012), p. e42095.
- [40] F. CIOCCHETTA AND J. HILLSTON, *Bio-pepa : A framework for the modelling and analysis of biological systems*, Theoretical Computer Science, 410 (2009), pp. 3065–3084.
- [41] E. M. CLARKE, E. A. EMERSON, AND A. P. SISTLA, *Automatic verification of finite-state concurrent systems using temporal logic specifications*, ACM Transactions on Programming Languages and Systems, 8 (1986), pp. 244–263.
- [42] F. S. COLLINS, M. MORGAN, AND A. PATRINOS, *The human genome project : lessons from large-scale biology*, Science, 300 (2003), pp. 286–290.
- [43] J.-P. COMET, G. BERNOT, A. DAS, F. DIENER, C. MASSOT, AND A. CESSIEUX, *Simplified models for the mammalian circadian clock*, Procedia Computer Science, 11 (2012), pp. 127–138.
- [44] J.-P. COMET, J. FROMENTIN, G. BERNOT, AND O. ROUX, *A formal model for gene regulatory networks with time delays*, in Computational Systems-Biology and Bioinformatics, Springer, 2010, pp. 1–13.

- [45] F. CORBLIN, E. FANCHON, AND L. TRILLING, *Applications of a formal approach to decipher discrete genetic networks*, BMC Bioinformatics, 11 (2010), p. 385.
- [46] F. CORBLIN, E. FANCHON, L. TRILLING, C. CHAOUIYA, AND D. THIEFFRY, *Automatic inference of regulatory and dynamical properties from incomplete gene interaction and expression data*, in IPCAT, 2012, pp. 25–30.
- [47] P. COUSOT AND R. COUSOT, *Abstract interpretation : A unified lattice model for static analysis of programs by construction or approximation of fixpoints*, in Proceedings of the 4th ACM SIGACT-SIGPLAN Symposium on Principles of Programming Languages (POPL'77), New York, NY, USA, 1977, ACM, pp. 238–252.
- [48] V. DANOS, J. FERET, W. FONTANA, R. HARMER, AND J. KRIVINE, *Rule-based modelling of cellular signalling*, in CONCUR 2007–Concurrency Theory, Springer, 2007, pp. 17–41.
- [49] V. DANOS, J. FERET, W. FONTANA, AND J. KRIVINE, *Abstract interpretation of cellular signalling networks*, in Verification, Model Checking, and Abstract Interpretation, Springer, 2008, pp. 83–97.
- [50] A. DAVID, G. BEHRMANN, P. BULYCHEV, J. BYG, T. CHATAIN, K. LARSEN, P. PETTERSSON, J. RASMUSSEN, J. SRBA, W. YI, K. JOERGENSEN, D. LIME, M. MAGNIN, O. ROUX, AND L.-M. TRAONOUZ, *Communicating Embedded Systems - Software and Design*, ISTE Publishing/John Wiley, 2009, ch. Tools for Model-Checking Timed Systems.
- [51] A. DAVID, G. BEHRMANN, K. LARSEN, P. PETTERSSON, J. RASMUSSEN, W. YI, D. LIME, M. MAGNIN, AND O. ROUX, *Approches formelles des systèmes embarqués communicants*, Hermes Lavoisier, 2008, ch. Outils de Model Checking.
- [52] H. DE JONG, *Modeling and simulation of genetic regulatory systems : a literature review*, Journal of computational biology, 9 (2002), pp. 67–103.
- [53] E. DE MARIA, F. FAGES, AND S. SOLIMAN, *On coupling models using model-checking : Effects of irinotecan injections on the mammalian cell cycle*, in Computational Methods in Systems Biology, Springer, 2009, pp. 142–157.
- [54] J. DEMONGEOT, J. ARACENA, F. THUDEROZ, T.-P. BAUM, AND O. COHEN, *Genetic regulation networks : circuits, regulons and attractors*, Comptes Rendus Biologies, 326 (2003), pp. 171–188.
- [55] C. DI GIUSTO, H. KLAUDEL, AND F. DELAPLACE, *Systemic approach for toxicity analysis*, in 5th International Workshop on Biological Processes and Petri Nets (BioPPN 2014), vol. 1159, 2014, pp. 30–44.

- [56] M. DURZINSKY, W. MARWAN, M. OSTROWSKI, T. SCHAUB, AND A. WAGLER, *Automatic network reconstruction using asp*, Theory and Practice of Logic Programming, 11 (2011), pp. 749–766.
- [57] J.-P. ELLOY AND M. MAGNIN, *Le numérique : une bibliothèque universelle pour la création scientifique ?*, Revue de la BNU, 10 (2014), pp. 68–75.
- [58] C. EVAIN, S. CAROLAN, AND M. MAGNIN, *Promoting science and engineering to generation z*, in ASME 2012 11th Biennial Conference on Engineering Systems Design and Analysis, American Society of Mechanical Engineers, 2012.
- [59] C. EVAIN AND C. DE MARCO, *Reinventing the book : The ezoombook platform*, Ubiquitous Learning : An International Journal, 7 (2014), pp. 1–13.
- [60] C. EVAIN, M. MAGNIN, AND G. MOREAU, *Diffuser les bonnes pratiques autour de l’usage des blogs en pédagogie*, in 7ème Colloque International Technologies de l’Information et de la Communication pour l’Enseignement (TICE 2010). Retour d’Expérience (ReX), 2010.
- [61] F. FAGES AND S. SOLIMAN, *Formal cell biology in biocham*, in Formal Methods for Computational Systems Biology, Springer, 2008, pp. 54–80.
- [62] F. FAGES, S. SOLIMAN, AND N. CHABRIER-RIVIER, *Modelling and querying interaction networks in the biochemical abstract machine biocham*, Journal of Biological Physics and Chemistry, 4 (2004), pp. 64–73.
- [63] A. FAURÉ, A. NALDI, C. CHAOUYYA, AND D. THIEFFRY, *Dynamical analysis of a generic boolean model for the control of the mammalian cell cycle*, Bioinformatics, 22 (2006), pp. e124–e131.
- [64] M. FOLSCHETTE, *Modélisation algébrique de la dynamique multi-échelles des réseaux de régulation biologique*, PhD thesis, University of Nantes, ED STIM, Ecole Centrale de Nantes, Université de Nantes, Nantes, Oct. 2014.  
Sous la direction de : Olivier Roux et Morgan Magnin. Jury : Mireille Régnier (prés.), Jean-Paul Comet (rapp.), Anne Siegel (rapp.), Denis Thieffry.
- [65] M. FOLSCHETTE, L. PAULEVÉ, K. INOUE, M. MAGNIN, AND O. ROUX, *Abducing biological regulatory networks from process hitting models*, in ECML-PKDD 2012 Workshop on Learning and Discovery in Symbolic Systems Biology, 2012, pp. 24–35.
- [66] M. FOLSCHETTE, L. PAULEVÉ, K. INOUE, M. MAGNIN, AND O. ROUX, *Concretizing the process hitting into biological regulatory networks*, in Computational Methods in Systems Biology, D. Gilbert and M. Heiner, eds., Lecture Notes in Computer Science, Springer, 2012, pp. 166–186.

- 
- [67] M. FOLSCHETTE, L. PAULEVÉ, K. INOUE, M. MAGNIN, AND O. ROUX, *Identification of biological regulatory networks from process hitting models*, Theoretical Computer Science, 568 (2015), pp. 49–71.
- [68] M. FOLSCHETTE, L. PAULEVÉ, M. MAGNIN, AND O. ROUX, *Under-approximation of reachability in multivalued asynchronous networks*, Electronic Notes in Theoretical Computer Science, 299 (2013), pp. 33–51.  
Proceedings of the fourth International Workshop on Interactions between Computer Science and Biology (CS2Bio'13).
- [69] M. FOLSCHETTE, L. PAULEVÉ, M. MAGNIN, AND O. ROUX, *Sufficient conditions for reachability in automata networks with priorities*, Theoretical Computer Science, 608 (2015), pp. 66–83.
- [70] P. FRANÇOIS, V. HAKIM, AND E. D. SIGGIA, *Deriving structure from evolution : metazoan segmentation*, Molecular Systems Biology, 3 (2007).
- [71] J. FROMENTIN, J.-P. COMET, P. LE GALL, AND O. ROUX, *Analysing gene regulatory networks by both constraint programming and model-checking*, in EMBC'07, 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Lyon, France, 2007, August 23-26, pp. 4595–4598.
- [72] G. GARDEY, D. LIME, M. MAGNIN, AND O. ROUX, *Romeo : A tool for analyzing time petri nets*, in Computer Aided Verification, K. Etessami and S. Rajamani, eds., vol. 3576 of Lecture Notes in Computer Science, Springer, 2005, pp. 418–423.
- [73] A. GARG, A. DI CARA, I. XENARIOS, L. MENDOZA, AND G. DE MICHELI, *Synchronous versus asynchronous modeling of gene regulatory networks*, Bioinformatics, 24 (2008), pp. 1917–1925.
- [74] F. GHANBARNEJAD, *Perturbations in boolean networks*, arXiv preprint arXiv :1111.5334, (2011).
- [75] J.-M. GILLIOT, A.-C. GROLLEAU, M. MAGNIN, AND C. VAUFREY, *Itypa, un premier mooc francophone et connectiviste*, in QPES 2013 : colloque questions de pédagogies dans l'enseignement supérieur, 2013.
- [76] V. GRIMM AND J. M. CALABRESE, *What is resilience ? a short introduction*, in Viability and Resilience of Complex Systems, Springer, 2011, pp. 3–13.
- [77] C. GUZIOŁOWSKI, S. VIDELA, F. EDUATI, S. THIELE, T. COKELAER, A. SIEGEL, AND J. SAEZ-RODRIGUEZ, *Exhaustively characterizing feasible logic models of a signaling network using answer set programming*, Bioinformatics, (2013), p. btt393.



- [78] S. HADDAD, L. MOKDAD, AND P. MOREAUX, *A new approach to the evaluation of non Markovian stochastic Petri nets*, in Proceedings of the 27th International Conference on Applications and Theory of Petri Nets, vol. 4024 of Lecture Notes in Computer Science, Springer, 2006, pp. 221–240.
- [79] A. HAMEZ, Y. THIERRY-MIEG, AND F. KORDON, *Building efficient model checkers using hierarchical set decision diagrams and automatic saturation*, Fundamenta Informaticae, 94 (2009), pp. 413–437.
- [80] H. HARA, *A relation between non-markov and markov processes*, Zeitschrift für Physik B Condensed Matter, 36 (1980), pp. 369–372.
- [81] M. HEINER, D. GILBERT, AND R. DONALDSON, *Petri nets for systems and synthetic biology*, in Proceedings of the Formal Methods for the Design of Computer, Communication, and Software Systems 8th International Conference on Formal Methods for Computational Systems Biology, SFM’08, Berlin, Heidelberg, 2008, Springer-Verlag, pp. 215–264.
- [82] C. S. HOLLING, *Resilience and stability of ecological systems*, Annual review of ecology and systematics, (1973), pp. 1–23.
- [83] T. IDEKER, T. GALITSKI, AND L. HOOD, *A new approach to decoding life : systems biology*, Annual review of genomics and human genetics, 2 (2001), pp. 343–372.
- [84] K. INOUE, *Logic programming for boolean networks*, in Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence - Volume Volume Two, IJCAI’11, AAAI Press, 2011, pp. 924–930.
- [85] ———, *Logic programming for boolean networks*, in Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence, IJCAI’11, AAAI Press, 2011, pp. 924–930.
- [86] K. INOUE, A. DONCESCU, AND H. NABESHIMA, *Completing causal networks by meta-level abduction*, Machine learning, 91 (2013), pp. 239–277.
- [87] K. INOUE, T. RIBEIRO, AND C. SAKAMA, *Learning from interpretation transition*, Machine learning, 94 (2014), pp. 51–79.
- [88] K. INOUE AND C. SAKAMA, *Oscillating behavior of logic programs*, in Correct Reasoning, Springer, 2012, pp. 345–362.
- [89] M. JOHN, C. LHOSSAINE, J. NIEHREN, AND A. M. UHRMACHER, *The attributed pi-calculus with priorities*, in Transactions on Computational Systems Biology XII, Springer, 2010, pp. 13–76.

- [90] S. A. KAUFFMAN, *Metabolic stability and epigenesis in randomly constructed genetic nets*, Journal of theoretical biology, 22 (1969), pp. 437–467.
- [91] Z. KHALIS, J.-P. COMET, A. RICHARD, AND G. BERNOT, *The SMBioNet method for discovering models of gene regulatory networks*, Genes, Genomes and Genomics, 3(special issue 1) (2009), pp. 15–22.
- [92] H. KITANO, *Computational systems biology*, Nature, 420 (2002), pp. 206–210.
- [93] ———, *Systems biology : a brief overview*, Science, 295 (2002), pp. 1662–1664.
- [94] S. KLAMT, J. SAEZ-RODRIGUEZ, J. LINDQUIST, L. SIMEONI, AND E. GILLES, *A methodology for the structural and functional analysis of signaling and regulatory networks*, BMC Bioinformatics, 7 (2006), p. 56.
- [95] I. KOCH AND M. HEINER, *Petri Nets*, Wiley Book Series on Bioinformatik, John Wiley & Sons, 2008, ch. 7, pp. 139–179.  
Series Eds. Yi Pan, Albert Y. Zomaya.
- [96] C. KOH, F.-X. WU, G. SELVARAJ, AND A. J. KUSALIK, *Using a state-space model and location analysis to infer time-delayed regulatory networks*, EURASIP Journal on Bioinformatics and Systems Biology, 2009 (2009), p. 14.
- [97] A. S. KOKSAL, Y. PU, S. SRIVASTAVA, R. BODIK, J. FISHER, AND N. PITERMAN, *Synthesis of biological models from mutation experiments*, ACM SIGPLAN Notices, 48 (2013), pp. 469–482.
- [98] C. KUTTLER, C. LHOSSAINE, AND J. NIEHREN, *A stochastic pi-calculus for concurrent objects*, Springer, 2007.
- [99] C. KUTTLER AND J. NIEHREN, *Gene regulation in the pi calculus : Simulating cooperativity at the lambda switch*, in Transactions on computational systems biology VII, Springer, 2006, pp. 24–55.
- [100] M. KWIATKOWSKA, G. NORMAN, AND D. PARKER, *Prism : Probabilistic symbolic model checker*, in Computer performance evaluation : modelling techniques and tools, Springer, 2002, pp. 200–204.
- [101] X. LI, S. RAO, W. JIANG, C. LI, Y. XIAO, Z. GUO, Q. ZHANG, L. WANG, L. DU, J. LI, ET AL., *Discovery of time-delayed gene regulatory networks based on temporal gene expression profiling*, BMC bioinformatics, 7 (2006), p. 26.
- [102] T.-F. LIU, W.-K. SUNG, AND A. MITTAL, *Learning multi-time delay gene network using bayesian network framework*, in Tools with Artificial Intelligence, 2004. ICTAI 2004. 16th IEEE International Conference on, IEEE, 2004, pp. 640–645.

- [103] M. LOPES AND G. BONTEMPI, *Experimental assessment of static and dynamic algorithms for gene regulation inference from time series expression data.*, Frontiers in genetics, 4 (2013), p. 303.
- [104] M. MAGNIN, *Réseaux de Petri à chronomètres : temps dense et temps discret*, PhD thesis, University of Nantes, ED STIM, Ecole Centrale de Nantes, Université de Nantes, Nantes, Dec. 2007.  
Sous la direction de : Olivier H. Roux et Pierre Molinaro. Jury : Béatrice Bérard (prés.), Guy Juanole (rapp.), Patrice Moreaux (rapp.), Hanifa Boucheneb, Jean-Pierre Elloy (invité).
- [105] M. MAGNIN, *Les interfaces entre le manga et la société japonaise : renforcer l'utopie pour affronter la crise*, in 4èmes Rencontres Internationales Jules Verne, 2012.
- [106] M. MAGNIN, *Relever le défi des moocs*, Arabesques, 71 (2013), pp. 20–21.
- [107] M. MAGNIN, D. LIME, AND O. H. ROUX, *An efficient method for computing exact state space of Petri nets with stopwatches*, in third International Workshop on Software Model-Checking (SoftMC'05), vol. 144 of Electronic Notes in Theoretical Computer Science, Edinburgh, Scotland, UK, July 2005, Elsevier, pp. 59–77.
- [108] M. MAGNIN, D. LIME, AND O. H. ROUX, *Symbolic state space of stopwatch petri nets with discrete-time semantics*, in The 29th International Conference on Application and Theory of Petri Nets and other models of concurrency (ICATPN 2008), K. van Hee and R. Valk, eds., vol. 5062 of Lecture Notes in Computer Science, Xi'an, China, June 2008, Springer, pp. 307–326.
- [109] M. MAGNIN, P. MOLINARO, AND O. H. ROUX, *Decidability, expressivity and state-space computation of stopwatch petri nets with discrete-time semantics*, in 8th International Workshop on Discrete Event Systems, IEEE, 2006, pp. 33–38.
- [110] M. MAGNIN, P. MOLINARO, AND O. H. ROUX, *Expressiveness of petri nets with stopwatches. dense-time part*, Fundamenta Informaticae, 97 (2009), pp. 111–138.
- [111] M. MAGNIN, P. MOLINARO, AND O. H. ROUX, *Expressiveness of petri nets with stopwatches. discrete-time part*, Fundamenta Informaticae, 97 (2009), pp. 139–176.
- [112] M. MAGNIN AND G. MOREAU, *Retour sur l'usage de tablet pc à l'école centrale de nantes : autonomie et initiatives*, in 7ème Colloque International Technologies de l'Information et de la Communication pour l'Enseignement (TICE 2010). Retour d'Expérience (ReX), 2010.

- 
- [113] M. MAGNIN AND G. MOREAU, *Utilisation des tests de concordance de scripts pour l'évaluation en informatique*, in 7ème Colloque International Technologies de l'Information et de la Communication pour l'Enseignement (TICE'2010). Communication scientifique (ComSci), 2010.
  - [114] M. MAGNIN AND G. MOREAU, *La professionnalisation des étudiants en informatique par les projets libres*, in 6ème Colloque Questions de pédagogies dans l'enseignement supérieur (QPES), 2011.
  - [115] M. MAGNIN AND G. MOREAU, *Utilisation des tests de concordance de scripts pour l'évaluation en informatique*, Revue Sticef.org, 9 (2011), p. 01.
  - [116] M. MAGNIN, G. MOREAU, V. TOURRE, N. VAROQUAUX, AND B. VIALLE, *Apports de l'application web libre markus dans le processus de correction de travaux en informatique*, in 8ème Colloque International Technologies de l'Information et de la Communication pour l'Enseignement (TICE 2012). Retour d'Expérience (ReX), 2012.
  - [117] M. MAGNIN, G. MOREAU, N. VAROQUAUX, B. VIALLE, K. REID, M. CONLEY, AND S. GEHWOLF, *Markus : An open-source web application to annotate student papers on-line*, in ASME 2012 11th Biennial Conference on Engineering Systems Design and Analysis, American Society of Mechanical Engineers, 2012, pp. 301–307.
  - [118] S. MANGAN AND U. ALON, *Structure and function of the feed-forward loop network motif*, Proceedings of the National Academy of Sciences, 100 (2003), pp. 11980–11985.
  - [119] V. MARX, *Biology : The big challenges of big data*, Nature, 498 (2013), pp. 255–260.
  - [120] H. MATSUNO, A. DOI, M. NAGASAKI, AND S. MIYANO, *Hybrid petri net representation of gene regulatory network*, in Pacific Symposium on Biocomputing, vol. 5, World Scientific Press Singapore, 2000, p. 87.
  - [121] M. MAURIN, M. MAGNIN, AND O. ROUX, *Modeling of genetic regulatory network in stochastic  $\pi$ -calculus*, in Bioinformatics and Computational Biology, Springer, 2009, pp. 282–294.
  - [122] A. A. MELKMAN, T. TAMURA, AND T. AKUTSU, *Determining a singleton attractor of an and/or boolean network in  $o(n \cdot 587)$  time*, Information Processing Letters, 110 (2010), pp. 565–569.
  - [123] R. MILNER, *Communication and Concurrency*, Prentice-Hall, Inc., 1989.
  - [124] G. MOREAU, M. SERVIÈRES, J.-M. NORMAND, AND M. MAGNIN, *Challenges of image-based crowd-sourcing for situation awareness in disaster management*, in Improving Disaster Resilience and Mitigation-IT Means and Tools, Springer, 2014, pp. 103–118.

- [125] N. NAKAJIMA AND T. AKUTSU, *Network completion for time varying genetic networks*, in Complex, Intelligent, and Software Intensive Systems (CISIS), 2013 Seventh International Conference on, IEEE, 2013, pp. 553–558.
- [126] —, *Exact and heuristic methods for network completion for time-varying genetic networks*, BioMed research international, 2014 (2014).
- [127] —, *Network completion for static gene expression data*, Advances in bioinformatics, 2014 (2014).
- [128] A. NALDI, D. THIEFFRY, AND C. CHAOUIYA, *Decision diagrams for the representation and analysis of logical models of genetic networks*, in Computational Methods in Systems Biology, Springer, 2007, pp. 233–247.
- [129] I. NIEMELÄ, *Logic programs with stable model semantics as a constraint programming paradigm*, Ann. Math. Artif. Intell., 25 (1999), pp. 241–273.
- [130] M. NOUAL, *Synchronism vs asynchronism in boolean networks*, CoRR, abs/1104.4039 (2011).
- [131] M. NOUAL, D. REGNAULT, AND S. SENÉ, *Non-monotony and boolean automata networks*, CoRR, abs/1111.4552 (2011).
- [132] E. P. ODUM, H. T. ODUM, AND J. ANDREWS, *Fundamentals of ecology*, vol. 3, Saunders Philadelphia, 1971.
- [133] L. PAULEVÉ, *Modelling, Simulation and Verification of Large Biological Regulatory Networks*, thèse, Ecole Centrale de Nantes (ECN), Oct. 2011.
- [134] L. PAULEVÉ, G. ANDRIEUX, AND H. KOEPL, *Under-approximating cut sets for reachability in large scale automata networks*, in Computer Aided Verification, N. Sharygina and H. Veith, eds., vol. 8044 of Lecture Notes in Computer Science, Springer Berlin Heidelberg, 2013, pp. 69–84.
- [135] L. PAULEVÉ, C. CHANCELLOR, M. FOLSCHETTE, M. MAGNIN, AND O. ROUX, *Logical Modeling of Biological Systems*, Wiley, 2014, ch. Analyzing Large Network Dynamics with Process Hitting, pp. 125 – 166.
- [136] L. PAULEVÉ, M. FOLSCHETTE, M. MAGNIN, AND O. ROUX, *Analyses statiques de la dynamique des réseaux d’automates indéterministes*, tech. rep., Mar. 2014.
- [137] L. PAULEVÉ, M. FOLSCHETTE, M. MAGNIN, AND O. ROUX, *Analyses statiques de la dynamique des réseaux d’automates indéterministes*, Revue des Sciences et Technologies de l’Information - Série TSI : Technique et Science Informatiques, 34 (2015), p. 22.

- [138] L. PAULEVÉ, M. MAGNIN, AND O. ROUX, *Abstract Interpretation of Dynamics of Biological Regulatory Networks*, Electronic Notes in Theoretical Computer Science, 272 (2011), pp. 43–56.  
 Proceedings of The First International Workshop on Static Analysis and Systems Biology (SASB 2010).
- [139] L. PAULEVÉ, M. MAGNIN, AND O. ROUX, *Refining dynamics of gene regulatory networks in a stochastic  $\pi$ -calculus framework*, in Transactions on computational systems biology XIII, Springer, 2011, pp. 171–191.
- [140] L. PAULEVÉ, M. MAGNIN, AND O. ROUX, *Tuning temporal features within the stochastic  $\pi$ -calculus*, Software Engineering, IEEE Transactions on, 37 (2011), pp. 858–871.
- [141] L. PAULEVÉ, M. MAGNIN, AND O. ROUX, *From the Process Hitting to Petri Nets and Back*, Technical Report hal-00744807, Oct. 2012.
- [142] L. PAULEVÉ, M. MAGNIN, AND O. ROUX, *Static analysis of biological regulatory networks dynamics using abstract interpretation*, Mathematical Structures in Computer Science, 22 (2012), pp. 651–685.
- [143] L. PAULEVÉ AND A. RICHARD, *Static analysis of boolean networks based on interaction graphs : a survey*, Electronic Notes in Theoretical Computer Science, 284 (2011), pp. 93 – 104.  
 Proceedings of The Second International Workshop on Static Analysis and Systems Biology (SASB 2011).
- [144] J. L. PETERSON, *Petri nets*, ACM Computing Surveys, 9 (1977), pp. 223–252.
- [145] C. A. PETRI, *Kommunikation mit Automaten*, PhD thesis, Fakultät für Mathematik und Physik, Technische Hochschule Darmstadt, Darmstadt (Allemagne), 1962.
- [146] A. PHILLIPS AND L. CARDELLI, *A programming language for composable dna circuits*, Journal of the Royal Society Interface, 6 (2009), pp. S419–S436.
- [147] A. PNUELI, *The temporal logic of programs*, in SFCS’77 : Proceedings of the 18th Annual Symposium on Foundations of Computer Science, Washington, DC, USA, 1977, IEEE Computer Society, pp. 46–57.
- [148] C. PRIAMI, *Stochastic  $\pi$ -calculus*, The Computer Journal, 38 (1995), pp. 578–589.
- [149] É. REMY, P. RUET, AND D. THIEFFRY, *Graphic requirements for multistability and attractive cycles in a boolean dynamical framework*, Advances in Applied Mathematics, 41 (2008), pp. 335 – 350.

- [150] T. RIBEIRO, M. MAGNIN, AND K. INOUE, *Learning delayed influence of dynamical systems from interpretation transition*, in The 24th International Conference on Inductive Logic Programming, 2014.  
Short paper (<http://tony.research.free.fr/paper/ILP2014short>).
- [151] T. RIBEIRO, M. MAGNIN, K. INOUE, AND C. SAKAMA, *Learning delayed influences of biological systems*, *Frontiers in Bioengineering and Biotechnology*, 2 (2015), p. 81.
- [152] A. RICHARD, *Negative circuits and sustained oscillations in asynchronous automata networks*, *Advances in Applied Mathematics*, 44 (2010), pp. 378–392.
- [153] A. RICHARD AND J.-P. COMET, *Necessary conditions for multistationarity in discrete dynamical systems*, *Discrete Applied Mathematics*, 155 (2007), pp. 2403 – 2413.
- [154] A. RICHARD, J.-P. COMET, AND G. BERNOT, *Formal methods for modeling biological regulatory networks*, in *Modern Formal Methods and Applications*, Springer, 2006, pp. 83–122.
- [155] —, *Modern Formal Methods and Applications*, Springer Netherlands, 2006, ch. Formal Methods for Modeling Biological Regulatory Networks, pp. 83–122.
- [156] J. SAEZ-RODRIGUEZ, L. SIMEONI, J. A. LINDQUIST, R. HEMENWAY, U. BOMMARDT, B. ARNDT, U.-U. HAUS, R. WEISMANTEL, E. D. GILLES, S. KLAMT, AND B. SCHRAVEN, *A logical model provides insights into t cell receptor signaling*, *PLoS Computational Biology*, 3 (2007), p. e163.
- [157] O. SAHIN, H. FROHLICH, C. LOBKE, U. KORF, S. BURMESTER, M. MAJETY, J. MATTERN, I. SCHUPP, C. CHAOUIYA, D. THIEFFRY, A. POUSTKA, S. WIEMANN, T. BEISSBARTH, AND D. ARLT, *Modeling ERBB receptor-regulated G1/S transition to find novel targets for de novo trastuzumab resistance*, *BMC Systems Biology*, 3 (2009).
- [158] R. SAMAGA, J. SAEZ-RODRIGUEZ, L. G. ALEXOPOULOS, P. K. SORGER, AND S. KLAMT, *The logic of egfr/erbb signaling : Theoretical properties and analysis of high-throughput data*, *PLoS Computational Biology*, 5 (2009), p. e1000438.
- [159] N. SCHWIND, M. MAGNIN, AND K. INOUE, *Resilience of event-driven dynamic systems*, in 27th Annual Conference on Japanese Society of Artificial Intelligence (JSAI 2013), vol. 27, 2013, pp. 1–8.
- [160] N. SCHWIND, M. MAGNIN, L. INOUE, T. OKIMOTO, T. SATO, K. MINAMI, AND H. MARYAMA, *Formalization of resilience for constraint-based dynamic systems*, *Journal of Reliable Intelligent Environments*, (2015), pp. 1–19.

- 
- [161] C. SEIDNER, J. LERAT, O. ROUX, AND M. MAGNIN, *Vérification dynamique et formelle d'un système décrit par son architecture fonctionnelle à l'aide de réseaux de petri temporels (tpn) : promesses et perspectives*, in 4e conférence annuelle d'ingénierie système (AFIS), 2006.
- [162] H. SIEBERT AND A. BOCKMAYR, *Temporal constraints in the logical analysis of regulatory networks*, Theoretical Computer Science, 391 (2008), pp. 258–275.
- [163] A. SILVESCU AND V. HONAVAR, *Temporal boolean network models of genetic networks and their inference from gene expression time series*, Complex Systems, 13 (2001), pp. 61–78.
- [164] L. A. SOINOV, M. A. KRESTYANINOVA, AND A. BRAZMA, *Towards reconstruction of gene networks from expression data by supervised learning*, Genome biology, 4 (2003), p. 6.
- [165] P. T. SPELLMAN, G. SHERLOCK, M. Q. ZHANG, V. R. IYER, K. ANDERS, M. B. EISEN, P. O. BROWN, D. BOTSTEIN, AND B. FUTCHER, *Comprehensive identification of cell cycle-regulated genes of the yeast saccharomyces cerevisiae by microarray hybridization*, Molecular biology of the cell, 9 (1998), pp. 3273–3297.
- [166] J. STELLING, S. KLAMT, K. BETTENBROCK, S. SCHUSTER, AND E. D. GILLES, *Metabolic network structure determines key aspects of functionality and regulation*, Nature, 420 (2002), pp. 190–193.
- [167] D. THIEFFRY AND R. THOMAS, *Dynamical behaviour of biological regulatory networks-ii. immunity control in bacteriophage lambda*, Bulletin of Mathematical Biology, 57 (1995), pp. 277–297.
- [168] R. THOMAS, *Regulatory networks seen as asynchronous automata : a logical description*, Journal of Theoretical Biology, 153 (1991), pp. 1–23.
- [169] J. UHLENDORF, A. MIERMONT, T. DELAVEAU, G. CHARVIN, F. FAGES, S. BOTTANI, G. BATT, AND P. HERSEN, *Long-term model predictive control of gene expression at the population and single-cell levels*, Proceedings of the National Academy of Sciences, 109 (2012), pp. 14271–14276.
- [170] S. VIDELA, C. GUZIOLOWSKI, F. EDUATI, S. THIELE, M. GEBSER, J. NICOLAS, J. SAEZ-RODRIGUEZ, T. SCHAUB, AND A. SIEGEL, *Learning boolean logic models of signaling networks with asp*, Theoretical Computer Science, (2014).
- [171] N. WIENER, *Cybernetics : Control and communication in the animal and the machine*, Wiley New York, 1948.
- [172] Y. YAMAMOTO, A. ROUGNY, H. NABESHIMA, K. INOUE, H. MORIYA, C. FROIDEVAUX, AND K. IWANUMA, *Completing sbgn-af networks by logic-based hypothesis finding*, in Formal Methods in Macro-Biology, Springer, 2014, pp. 165–179.



## BIBLIOGRAPHIE

---

- [173] Z.-Y. ZHANG, *Time Series Segmentation for Gene Regulatory Process with Time-Window-Extension Technique*, (2008), pp. 198–203.



# Habilitation à Diriger des Recherches

**Morgan MAGNIN**

**Contribution à l'élaboration de connaissances qualitatives en bio-informatique**

**Contribution to the elaboration of qualitative knowledge in bioinformatics**

## Résumé

Ce manuscrit est un recueil de plusieurs travaux dans le domaine des méthodes formelles pour la biologie. Devant l'enjeu que représente la modélisation et l'analyse de la dynamique de réseaux de régulation biologiques à grande échelle, nous avons identifié une classe pertinente de modèles formels sur laquelle il est possible de mener des analyses efficaces de la dynamique. Après avoir discuté les différents critères de modélisation à prendre en compte en biologie, nous introduisons ainsi le formalisme des Frappes de Processus. Nous présentons ensuite les méthodes d'analyse conçues pour ce paradigme, leurs mérites et leurs limites. Enfin, nous revenons sur des résultats plus récents, consécutifs à l'établissement de liens fructueux entre l'apprentissage automatique, la programmation logique, le model-checking et la bio-informatique, ce qui nous permet de faire émerger un ensemble de nouvelles questions scientifiques.

## Mots clés

réseaux de régulation biologique, modélisation, frappes de processus, analyse de la dynamique, apprentissage automatique, vérification formelles, inférence de paramètres

## Abstract

This manuscript gathers several works in the field of formal methods for biology. In our works, we tackled the challenge of modeling and analyzing the dynamics of large-scale biological regulatory networks. To address this issue, we identified some relevant class of formal models on which it is possible to perform effective analysis of its dynamics. After discussing the different modeling criteria to be taken into account in biology, we introduce the Process Hitting framework. We then present the methods that we designed to analyze such models, and their respective merits and limitats. Finally, we give an overview of recent research aiming to build a fruitful link between machine learning, logic programming, model-checking and bioinformatics. This allows us to bring out a new set of scientific questions.

## Keywords

biological regulatory networks, modeling, process hitting, analysis of the dynamics, machine learning, model-checking, parameters inference